美国数学会经典影印系列

AMS
AMERICAN MATHEMATICAL SOCIETY

# Modern Geometric Structures and Fields

## 现代几何结构和场论

S. P. Novikov
I. A. Taimanov
Translated by Dmitry Chibisov

高等教育出版社

本书是作为微分流形上的几何和其上最重要结构的黎曼几何现代形式的一个引论。作者的观点是：黎曼几何所有构造的源头是一个流形，它使我们可以计算切向量的标量积。按此方式，作者向大家展示了黎曼几何对于现代数学的许多基本领域及其应用所产生的巨大影响。

- 几何是纯数学和自然科学——首先是物理学——之间的一个桥梁。自然界的基本规律就是由描述各种物理量的几何场之间的关系所构建的。
- 对几何对象整体性质的研究促进了拓扑学的意义深远的发展，这包括了纤维丛的拓扑与几何。
- 描述许多物理现象的哈密顿系统的几何理论推动了辛几何和泊松几何的发展。书中讲述的场论和多维变分学将数学与理论物理统一起来。
- 复几何与代数流形的几何则将黎曼几何与现代复分析，以及与代数和数论统一起来。

本书的预备知识包括几个基本的本科课程，诸如高等微积分、线性代数、常微分方程以及基础拓扑学。

本版只限于中华人民共和国境内发行。本版经由美国数学会授权仅在中华人民共和国境内销售，不得出口。

美国数学会经典影印系列

AMS
AMERICAN MATHEMATICAL SOCIETY

此书面对的是数学和理论物理专业的研究生，但对于教师也同样非常有用……比通常的微分几何教材内容更广。

—*EMS Newsletter*

此书包含丰富的一般理论、具体例题和代数计算。对于与物理学有内在联系的许多几何和代数领域，它是一本极具可读性的导引书，适合数学和理论物理专业的学生阅读。

—*Hubert Gollek for Zentralblatt MATH*

ISBN 978-7-04-046918-9

9 787040 469189 >

定价 269.00 元

美国数学会经典影印系列

AMS

AMERICAN MATHEMATICAL SOCIETY

# Modern Geometric Structures and Fields

## 现代几何结构和场论


S. P. Novikov

I. A. Taimanov

Translated by Dmitry Chibisov





高等教育出版社·北京

图字：01-2016-2513 号



**现代几何结构和场论**
Xiandai Jihe Jiegou he Changlun

**图书在版编目 (CIP) 数据**

现代几何结构和场论 = Modern Geometric Structures and Fields : 英文 / ( 俄罗斯 ) 诺维科夫 (S. P. Novikov) , ( 俄罗斯 ) 泰曼诺夫 (I. A. Taimanov) 著 . — 影印本 . — 北京 : 高等教育出版社 , 2018.8
ISBN 978-7-04-046918-9
Ⅰ. ①现… Ⅱ. ①诺…②泰… Ⅲ. ①几何—研究—英文Ⅳ. ①O18
中国版本图书馆 CIP 数据核字 (2016) 第 280455 号

策划编辑　李华英　　责任编辑　李华英
封面设计　张申申　　责任印制　刘思涵

出版发行 高等教育出版社
社址 北京市西城区德外大街 4 号
邮政编码 100120
购书热线 010-58581118
咨询电话 400-810-0598
网址 http://www.hep.edu.cn
http://www.hep.com.cn
网上订购 http://www.hepmall.com.cn
http://www.hepmall.com
http://www.hepmall.cn
印刷 山东临沂新华印刷物流集团

开本 787mm × 1092mm 1/16
印张 41.25
字数 1000 千字
版次 2018 年 8 月第 1 版
印次 2018 年 8 月第 1 次印刷
定价 269.00 元



[物 料 号 46918-00]

# 出版者的话

近年来，我国的科学技术取得了长足进步，特别是在数学等自然科学基础领域不断涌现出一流的研究成果。与此同时，国内的科研队伍与国外的交流合作也越来越密切，越来越多的科研工作者可以熟练地阅读英文文献，并在国际顶级期刊发表英文学术文章，在国外出版社出版英文学术著作。

然而，在国内阅读海外原版英文图书仍不是非常便捷。一方面，这些原版图书主要集中在科技、教育比较发达的大中城市的大型综合图书馆以及科研院所的资料室中，普通读者借阅不甚容易；另一方面，原版书价格昂贵，动辄上百美元，购买也很不方便。这极大地限制了科技工作者对于国外先进科学技术知识的获取，间接阻碍了我国科技的发展。

高等教育出版社本着植根教育、弘扬学术的宗旨服务我国广大科技和教育工作者，同美国数学会（American Mathematical Society）合作，在征求海内外众多专家学者意见的基础上，精选该学会近年出版的数十种专业著作，组织出版了“美国数学会经典影印系列”丛书。美国数学会创建于 1888 年，是国际上极具影响力的专业学术组织，目前拥有近 30000 会员和 580 余个机构成员，出版图书 3500 多种，冯·诺依曼、莱夫谢茨、陶哲轩等世界级数学大家都是其作者。本影印系列涵盖了代数、几何、分析、方程、拓扑、概率、动力系统等所有主要数学分支以及新近发展的数学主题。

我们希望这套书的出版，能够对国内的科研工作者、教育工作者以及青年学生起到重要的学术引领作用，也希望今后能有更多的海外优秀英文著作被介绍到中国。

高 等 教 育 出 版 社

2016 年 12 月

# Preface to the English Edition

The book presents basics of Riemannian geometry in its modern form as the geometry of differentiable manifolds and the most important structures on them. Let us recall that Riemannian geometry is built on the following fundamental idea: *The source of all geometric constructions is not the length but the inner product of two vectors.* This inner product can be symmetric and positive definite (classical Riemannian geometry) or symmetric indefinite (pseudo-Riemannian geometry and, in particular, the Lorentz geometry, important in physics). Moreover, this inner product can be complex Hermitian and even skew-symmetric (symplectic geometry). The modern point of view to any geometry also includes global topological aspects related to the corresponding structures. Therefore, in the book we pay attention to necessary basics of topology.

With this approach, Riemannian geometry has great influence on a number of fundamental areas of modern mathematics and its applications.

1. Geometry is a bridge between pure mathematics and natural sciences, first of all physics. Fundamental laws of nature are expressed as relations between geometric fields describing physical quantities.

2. The study of global properties of geometric objects leads to the far-reaching developments in topology, including topology and geometry of fiber bundles.

3. Geometric theory of Hamiltonian systems, which describe many physical phenomena, leads to the development of symplectic and Poisson geometry.

4. Geometry of complex and algebraic manifolds unifies Riemannian geometry with the modern complex analysis, as well as with algebra and number theory.

We tried, as much as possible, to present these ideas in our book. On the other hand, combinatorial geometry, which is based on a completely different set of ideas, remains beyond the scope of this book.

Now we briefly describe the contents of the book.

Chapters 1 and 2 are devoted to basics of various geometries in linear spaces: Euclidean, pseudo-Euclidean, and symplectic. Special attention is paid to the geometry of the Minkowski space, which is a necessary ingredient of relativity.

Chapters 3 and 4 are devoted to the geometry of two-dimensional manifolds; here we also present necessary facts of complex analysis. Some famous completely integrable systems of modern mathematical physics also appear in these chapters.

Basics of topology of smooth manifolds are presented in Chapter 5.

Some material about Lie groups is included in Chapter 6. In addition to classical theory of Lie groups and Lie algebras, here we pay attention to the theory of crystallographic groups and their modern generalizations related to the discovery of so-called quasicrystals.

Chapter 7 and 8 are devoted to classical tensor algebra and tensor calculus, whereas in Chapter 9 the theory of differential forms (antisymmetric tensors) is presented. It should be noted that already in Chapters 7 and 9 we encounter the so-called Grassmann (anticommuting) variables and the corresponding integration.

Chapter 10 contains the Riemannian theory of connections and curvature. The idea of a general connection in a bundle (Yang–Mills fields) also appears here.

Conformal geometry and complex geometry are the main subjects of Chapter 11.

In Chapter 12, we present elements of finite-dimensional Morse theory, the calculus of variations, and the theory of Hamiltonian systems. Geometric aspects of the theory of Hamiltonian systems are developed in Chapter 13, where the theory of Lagrangian and Poisson manifolds is introduced.

Chapter 14 contains elements of multidimensional calculus of variations. We should emphasize that, contrary to the majority of existing textbooks, the mathematical language of our exposition is compatible with the language of modern theoretical physics. In particular, we believe that such notions as the energy-momentum tensor, conservation laws, and the Noether theorem are an integral part of the exposition of this subject. Here we also touch

upon simplest topological aspects of the theory, such as the Morse index and harmonic forms on compact Riemannian manifolds from the point of view of Hodge theory.

In Chapter 15 we present the basic theory of the most important fields in physics such as the Einstein gravitation field, the Dirac spinor field, and the Yang–Mills fields associated to vector bundles. Here we also encounter topological phenomena such as the Chern and Pontryagin characteristic classes, and instantons.

Each chapter is supplemented by exercises that allow the reader to get a deeper understanding of the subject.

# Preface

As far back as in the late 1960s one of the authors of this book started preparations to writing a series of textbooks which would enable a modern young mathematician to learn geometry and topology. By that time, quite a number of problems of training nature were collected from teaching experience. These problems (mostly topological) were included into the textbooks [**DNF1**]–[**NF**] or published as a separate collection [**NMSF**]. The program mentioned above was substantially extended after we had looked at textbooks in theoretical physics (especially, the outstanding series by Landau and Lifshits, considerable part of which, e.g., the books [**LL1, LL2**], involve geometry in its modern sense), as well as from discussions with specialists in theoretical mechanics, especially L. I. Sedov and V. P. Myasnikov, in the Mechanics Division of the Mechanics and Mathematics Department of Moscow State University, who were extremely interested in establishing courses in modern geometry needed first of all in elasticity and other branches of mechanics. Remarkably, designing a modern course in geometry began in 1971 within the Mechanics, rather than the Mathematics Division of the department because this was where this knowledge was really needed. Mathematicians conceded to it later. Teaching these courses resulted in publication of lecture notes (in duplicated form):

S. P. Novikov, *Riemannian geometry and tensor analysis.* Parts I and II, Moscow State University, 1972/73.

Subsequently these courses were developed and extended, including, in particular, elements of topology, and were published as:

S. P. Novikov and A. T. Fomenko, *Riemannian geometry and tensor analysis.* Part III, Moscow State University, 1974.

After that, S. P. Novikov wrote the program of a course in the fundamentals of modern geometry and topology. It was realized in a series of books [**DNF1**]–[**NF**], written jointly with B. A. Dubrovin and A. T. Fomenko. Afterwards the topological part was completed by the book [**N1**], which contained a presentation of the basic ideas of classical topology as they have formed by the late 1960s–early 1970s. The later publication [**N2**] also included some recent advances in topology, but quite a number of deep new areas (such as, e.g., modern symplectic and contact topology, as well as new developments in the topology of 4-dimensional manifolds) were not covered yet. We recommend the book [**AN**]. We can definitely say that even now there is no comprehensible textbook that would cover the main achievements in the classical topology of the 1950s–1970s, to say nothing of the later period. Part II of the book [1] and the book [2] are insufficient; other books are sometimes unduly abstract; as a rule, they are devoted to special subjects and provide no systematic presentation of the progress made during this period, very important in the history of topology. Some well-written books (e.g., [**M1**]–[**MS**]) cover only particular areas of the theory. The book [**BT**] is a good supplement to [**DNF1, DNF2**], but its coverage is still insufficient.

Nevertheless, among our books, Part II of [**DNF1**] is a relatively good textbook containing a wide range of basic theory of differential topology in its interaction with physics. Nowadays this book could be modernized by essentially improving the technical level of presentation, but as a whole this book fulfills its task, together with the books [**DNF2**] and [**N1**], intended for a more sophisticated reader.

As for Part I, i.e., the basics of Riemannian geometry, it has become clear during the past 20 years that this book must be substantially revised, as far as the exposition of basics and more complete presentation of modern ideas are concerned. To this end, the courses [**T**] given by the second author, I. A. Taimanov, at Novosibirsk University proved to be useful. We joined our efforts in writing a new course using all the material mentioned above.

We believe that the time has come when a wide community of mathematicians working in geometry, analysis, and related fields will finally turn to the deep study of the contribution to mathematics made by theoretical physics of the 20th century. This turn was anticipated already 25 years ago, but its necessity was not realized then by a broad mathematical community. The advancements in this direction made in our books such as [**DNF1**] had not elicited a proper response among mathematicians for a long time. In our view, the situation is different nowadays. Mathematicians understand much better the necessity of studying mathematical tools used by physicists. Moreover, it appears that the state of the art in theoretical physics itself is

such that the deep mathematical methods created by physicists in the 20th century may be preserved for future mankind only by the mathematical community; we are concerned that the remarkable fusion of the strictly rational approach to the exploration of the real world with outstanding mathematical techniques may be lost.

Anyhow, we wrote this book for a broad readership of mathematicians and theoretical physicists. As before (see the Prefaces to the books [**DNF1**, **DNF2**]), we followed the principles that the book must be as comprehensible as possible and written with the minimal possible level of abstraction. A clear grasp of the natural essence of the subject must be achieved before starting its formalization. There is no sense to justify a theory which has not been perceived yet. The formal language disconnects rather than unites mathematics, and complicates rather than facilitates its understanding.

We hope that our ideas will be met with understanding by the mathematical community.

*The authors*

## Bibliography


[DNF1] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern geometry: Methods and applications.* Part I. *Geometry of surfaces, groups of transformations, and fields.* Part II. *Geometry and topology of manifolds.* 2nd ed., Nauka, Moscow, 1986; English transl. of 1st ed., *Modern geometry—Methods and applications.* Parts I and II, Graduate Texts in Mathematics, vols. 93 and 104, Springer-Verlag, New York, 1992.

[DNF2] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern geometry: Methods of the homology theory*, Nauka, Moscow, 1984; English transl. *Modern geometry—Methods and applications.* Part III, Graduate Texts in Mathematics, vol. 124, Springer-Verlag, New York, 1990.

[NF] S. P. Novikov and A. T. Fomenko, *Elements of differential geometry and topology*, Nauka, Moscow, 1987; English transl., Math. and Applications (Soviet series), vol. 60, Kluwer, Dordrecht, 1990.

[NMSF] S. P. Novikov, A. S. Mishchenko, Yu. P. Solov'ev, and A. T. Fomenko, *Problems in geometry: Differential geometry and topology*, Moscow State University Press, Moscow, 1978. (Russian)

[LL1] L. D. Landau and E. M. Lifshits, *Mechanics*, 3rd ed., Nauka, Moscow, 1973; English transl. of 1st ed., Pergamon Press, Oxford, and Addison-Wesley, Reading, MA, 1960.

[LL2] L. D. Landau and E. M. Lifshits, *The field theory*, 6th ed., Nauka, Moscow, 1973; English transl., Pergamon Press, Oxford-New York-Toronto, 1975.

[N1] S. P. Novikov, *Topology*-1, Current Problems in Mathematics. Fundamental Directions, VINITI, Moscow, 1986, vol. 12, pp. 5–252. (Russian)

[N2] S. P. Novikov, *Topology.* I, Encyclopaedia of Math. Sciences, Springer-Verlag, Berlin, 1996. (English transl. of [**N1**].)

[AN] V. I. Arnol'd and S. P. Novikov, eds., *Dynamical systems*, Current Problems in Mathematics. Fundamental Directions, VINITI, Moscow, 1985, vol. 4; English transl., Encyclopaedia of Math. Sciences, vol. 4, Springer-Verlag, Berlin, 2001.

[M1] J. W. Milnor, *Morse theory*, Ann. of Math. Studies, no. 51, Princeton Univ. Press, Princeton, NJ, 1963.

[M2] J. W. Milnor, *Lectures on the h-cobordism theorem*, Princeton Univ. Press, Princeton, NJ, 1965.

[A] M. F. Atiyah, *K-Theory*, Benjamin, New York, 1967.

[MS] D. McDuff and D. Salamon, *Introduction to symplectic topology*, Clarendon Press, Oxford, 1995.

[BT] R. Bott and L. W. Tu, *Differential forms in algebraic topology*, Springer-Verlag, New York-Heidelberg-Berlin, 1982.

[T] I. A. Taimanov, *Lectures on differential geometry*, Inst. Comp. Research, Moscow-Izhevsk, 2002. (Russian)

# Contents

Chapter 1

# Cartesian Spaces and Euclidean Geometry

## 1.1. Coordinates. Space-time

**1.1.1. Cartesian coordinates.** Geometry describes events in a space consisting of points $P, Q, \ldots$. We say that *Cartesian coordinates* are introduced in this space if each point of the space is associated with an ordered $n$-tuple $(x^1, \ldots, x^n)$ of real numbers, called the *coordinates of the point*, in such a way that the following conditions are satisfied:

1) every $n$-tuple $(x^1, \ldots, x^n)$ of real numbers corresponds to a point of the space, being the coordinates of this point;

2) the correspondence between the points of the space and the $n$-tuples of coordinates is one-to-one, i.e., the points with coordinates $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$ coincide if and only if $x^i = y^i$ for $i = 1, \ldots, n$.

A space where Cartesian coordinates $(x^1, \ldots, x^n)$ are introduced is called an *$n$-dimensional Cartesian space* and denoted by $\mathbb{R}^n$.

The number $n$ is called the *dimension* of the space.

Our physical space is of dimension $n = 3$, while time has dimension $n = 1$. According to the concepts of modern physics, space and time are inseparable, and one must consider the 4-dimensional *space-time continuum* whose points are instant events. This space is 4-dimensional, with coordinates the ordered quadruples $(t, x^1, x^2, x^3)$, where $t$ is the "time instant" when the event occurs, and $x^1, x^2, x^3$ are the coordinates of the "spacial

location" of the event. The 3-dimensional space of classical geometry is then the level surface $t =$ const, and the "life" of an object which may be regarded at any instant of time as a point (a so-called "point-particle") is identified with the *world line of the point-particle* $x^\alpha(t)$, $\alpha = 1, 2, 3$.

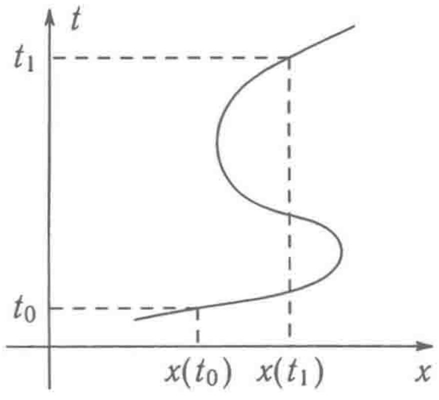


**Figure 1.1.** The world line of a particle.

We will often refer to the $n$-tuples $x = (x^1, \ldots, x^n)$ as the points of the Cartesian space $\mathbb{R}^n$.

**1.1.2. Change of coordinates.** A real-valued function $f(x)$ defined on an $n$-dimensional Cartesian space can be regarded as a function of $n$ real variables, $f(x) = f(x^1, \ldots, x^n)$ for $x = (x^1, \ldots, x^n)$.

In what follows we will also consider functions which will not be defined on the entire space $\mathbb{R}^n$, but rather on a part of $\mathbb{R}^n$, namely, on a domain of the space.

A *domain* (without boundary), or an *open set* in the space $\mathbb{R}^n$ is a set $U$ of points in $\mathbb{R}^n$ such that, together with a point $x$ belonging to $U$, all the points sufficiently close to $x$ also belong to $U$. More precisely, $U$ is a domain in $\mathbb{R}^n$ if for each point $x_0 = (x_0^1, \ldots, x_0^n)$ in $U$ there exists $\varepsilon > 0$ such that all the points $x = (x^1, \ldots, x^n)$ satisfying the inequalities

$$|x^i - x_0^i| < \varepsilon, \quad i = 1, \ldots, n,$$

lie in $U$. Any domain containing the point $x$ is said to be a *neighborhood* of this point.

A set $V \subset \mathbb{R}^n$ is *closed* if its complement $U = \mathbb{R}^n \setminus V$, i.e., the set consisting of all the points that do not lie in $V$, is open.

A function $f$ defined on the entire space $\mathbb{R}^n$ or on a domain in $\mathbb{R}^n$ is called *continuous* if it is continuous as a function of $n$ real variables that are Cartesian coordinates in $\mathbb{R}^n$. In a similar way, continuously differentiable (smooth) functions are defined.

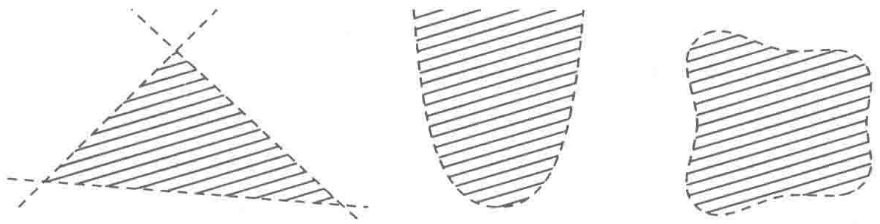

**Figure 1.2.** Examples of domains in a plane (a triangle bounded by straight lines, interior of a parabola, a bounded domain with smooth boundary).

EXAMPLE. Let $f_1, \ldots, f_m \colon \mathbb{R}^n \to \mathbb{R}$ be continuous real-valued functions on the space $\mathbb{R}^n$. Then the set $U$ of the points $x$ satisfying the inequalities

$$f_1(x) < 0, \ \ldots, \ f_m(x) < 0$$

is a domain in $\mathbb{R}^n$.

We prove this assertion. Let $x_0 = (x_0^1, \ldots, x_0^n)$ lie in $U$. Since the functions $f_1, \ldots, f_m$ are continuous, one can find positive numbers $\varepsilon_1, \ldots, \varepsilon_m$ such that for each $j$ the inequalities $|x^i - x_0^i| < \varepsilon_j$, $i = 1, \ldots, n$, imply the inequality $f_j(x^1, \ldots, x^n) < 0$. Putting $\varepsilon = \min_j \varepsilon_j$ we see that the set $U$ contains all the points such that $|x^i - x_0^i| < \varepsilon$, $i = 1, \ldots, n$. Therefore $U$ is a domain.

A domain $U$ is *bounded* if there is $R > 0$ such that all the points of $U$ satisfy the inequality

$$\sum_{i=1}^{n} (x^i)^2 \le R^2, \quad (x^1, \ldots, x^n) \in U.$$

Let $U$ and $V$ be domains in $\mathbb{R}^n$ and let $F \colon U \to V$ be a mapping specified coordinatewise by smooth functions

$$y^i = y^i(x^1, \ldots, x^n), \quad i = 1, \ldots, n$$

(we denote the coordinates in $U$ and $V$ by $x^i$, $i = 1, \ldots, n$, and $y^j$, $j = 1, \ldots, n$, respectively). The matrix

$$A(x_0) = (a_j^i) = \left(\frac{\partial y^i}{\partial x^j}\right)_{x^1 = x_0^1, \ldots, x^n = x_0^n}$$

is called the *Jacobi matrix* of the mapping $F$ at the point $x_0 = (x_0^1, \ldots, x_0^n)$ and is denoted by $\left(\frac{\partial y}{\partial x}\right)$. The *determinant* of this matrix is referred to as the *Jacobian* of the mapping $F$ and is denoted by $J$,

$$J = \det\left(\frac{\partial y}{\partial x}\right).$$

A point $x_0 = (x_0^1, \ldots, x_0^n)$ is said to be *nonsingular* for the mapping $F$ if the Jacobian at this point does not vanish.

The following inverse function theorem is a version of the implicit function theorem and is familiar to the reader from a calculus course.

**Theorem 1.1.** *If the mapping $F$ has a nonzero Jacobian at a point $x_0 = (x_0^1, \ldots, x_0^n)$ then in a sufficiently small neighborhood of the point $F(x_0) = (y_0^1, \ldots, y_0^n)$, where $y_0^i = y^i(x_0^1, \ldots, x_0^n)$, the coordinates $(x^1, \ldots, x^n)$ are uniquely determined by $y^1, \ldots, y^n$:*

$$x^i = x^i(y^1, \ldots, y^n), \quad i = 1, \ldots, n,$$

*and the Jacobi matrix $\left(\frac{\partial x}{\partial y}\right)$ of the inverse mapping $F^{-1}$ is the inverse of the matrix $\left(\frac{\partial y}{\partial x}\right)$:*

$$\frac{\partial x^i}{\partial y^j} \cdot \frac{\partial y^j}{\partial x^k} = \delta_k^i = \begin{cases} 1 & \text{for } i = k, \\ 0 & \text{for } i \neq k. \end{cases}$$

Here and elsewhere we apply the convention of summation over repeated indices, i.e., the presence of an index repeated once as a subscript and once as a superscript means summation over this index (over $j$ in the above expression).

This theorem implies that in a neighborhood of a nonsingular point the functions $y^1, \ldots, y^n$ specify new coordinates, which are connected with the coordinates $x^1, \ldots, x^n$ by smooth one-to-one transformations,

$$y^i = y^i(x^1, \ldots, x^n), \qquad x^j = x^j(y^1, \ldots, y^n), \quad 1 \leq i, j \leq n.$$

In general, whenever there exists such a mapping between the domains $U$ and $V$ in $\mathbb{R}^n$, we say that a change of coordinates is given.

The simplest example of a change of coordinates is provided by a linear transformation

$$y^i = a_j^i x^j, \quad i = 1, \ldots, n.$$

In this case the Jacobi matrix is constant and equal to $(a_j^i)$, while the inverse mapping has the form

$$x^i = b_j^i y^j, \quad i = 1, \ldots, n,$$

where $b_j^i a_k^j = \delta_k^i$.

However, in the plane and in three-dimensional space there are systems of coordinates, useful for applications, which are related to the Cartesian coordinates by transition functions $y^i = y^i(x^1, \ldots, x^n)$ nonlinear in $x^j$.

1. *Polar coordinates* $r, \varphi$ in the plane with Cartesian coordinates $x^1, x^2$:

$$x^1 = r \cos \varphi, \qquad x^2 = r \sin \varphi,$$

where $r \geq 0$. For an integer $k$, the pairs $(r, \varphi)$ and $(r, \varphi + 2\pi k)$ represent the same point; therefore, in order to make the coordinate $\varphi$ single-valued, we will impose the restriction $0 < \varphi < 2\pi$. The Jacobi matrix of the mapping $(r, \varphi) \to (x^1, x^2)$ is

$$\begin{pmatrix} \frac{\partial x^1}{\partial r} & \frac{\partial x^1}{\partial \varphi} \\ \frac{\partial x^2}{\partial r} & \frac{\partial x^2}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix},$$

and its determinant (Jacobian) equals

$$J = r \geq 0.$$

The Jacobian vanishes only for $r = 0$. Hence in the domain $r > 0$, $0 <$

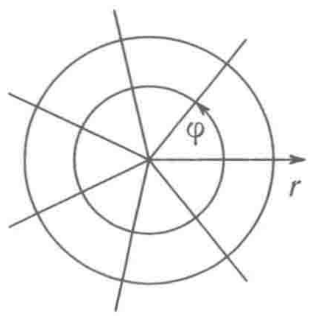


**Figure 1.3.** Polar coordinates.

$\varphi < 2\pi$, i.e., in the plane $\mathbb{R}^2$ with deleted half-line $x^1 \geq 0$, $x^2 = 0$, the coordinates $r, \varphi$ are uniquely determined and have no singular points.

2. *Cylindrical coordinates* $r, \varphi, z$ in $\mathbb{R}^3$:

$$x^1 = r\cos\varphi, \qquad x^2 = r\sin\varphi, \qquad x^3 = z,$$

where $x^1, x^2, x^3$ are Cartesian coordinates in $\mathbb{R}^3$. The Jacobian of the change of coordinates $(r, \varphi, z) \to (x^1, x^2, x^3)$ vanishes only for $r = 0$; therefore, this mapping has no singular points in the domain $r > 0$. As before, these coordinates are uniquely determined if $0 < \varphi < 2\pi$.

3. *Spherical coordinates* $r, \theta, \varphi$ in $\mathbb{R}^3$:

$$x^1 = r\cos\varphi\sin\theta, \qquad x^2 = r\sin\varphi\sin\theta, \qquad x^3 = r\cos\theta,$$

where $r \geq 0$, $0 \leq \theta \leq \pi$, $0 \leq \varphi \leq 2\pi$. The Jacobian equals $J = r^2\sin\theta$ and does not vanish in the domain $r > 0$, $\theta \neq 0, \pi$. In the domain $r > 0$, $0 < \theta < \pi$, $0 < \varphi < 2\pi$, the spherical coordinates are uniquely determined and the transition mapping $(r, \theta, \varphi) \to (x^1, x^2, x^3)$ has no singular points.

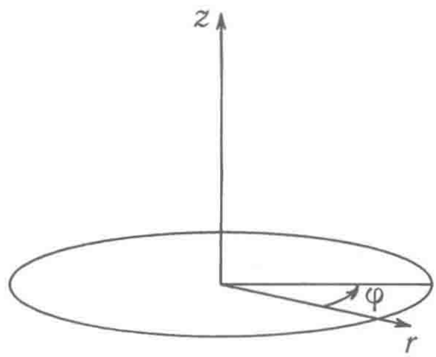


**Figure 1.4.** Cylindrical coordinates.

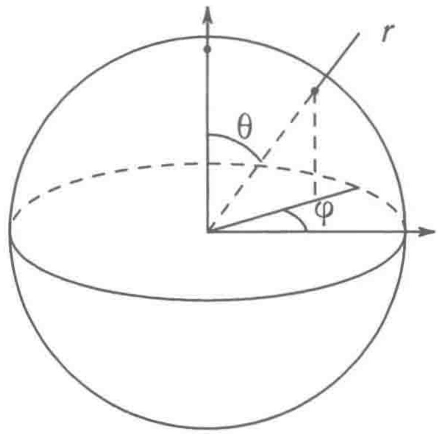


**Figure 1.5.** Spherical coordinates.

## 1.2. Euclidean geometry and linear algebra

**1.2.1. Vector spaces and scalar products.** The Cartesian space $\mathbb{R}^n$ is associated in a natural way with an $n$-dimensional *vector space*.

Namely, to each point $P = (x^1, \dots, x^n)$ corresponds the "radius-vector" of its coordinates, which is the $n$-tuple $(x^1, \dots, x^n)$ regarded as a vector $\xi$ in $n$-dimensional vector space. We will denote points and their radius-vectors by the same symbols: $x = (x^1, \dots, x^n)$.

The vectors may be added and multiplied by real numbers: if $\xi = (\xi^1, \dots, \xi^n)$ and $\eta = (\eta^1, \dots, \eta^n)$, then

$$\xi + \eta = (\xi^1 + \eta^1, \dots, \xi^n + \eta^n), \qquad \lambda\xi = (\lambda\xi^1, \dots, \lambda\xi^n).$$

The vectors $e_i = (0, \dots, 1, \dots, 0)$ (with 1 on the $i$th place) for $i = 1, \dots, n$ form a *basis* of the vector space.

Let $L$ be a subset of $\mathbb{R}^n$ specified by a system of linear equations

$$a^i_j x^j + b^i = 0, \quad i = 1, \dots, n - k$$

(with summation over repeated indices), or

$$\begin{cases} a_1^1x^1 + a_2^1x^2 + \cdots + a_n^1x^n + b^1 & = 0, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ a_1^{n-k}x^1 + a_2^{n-k}x^2 + \cdots + a_n^{n-k}x^n + b^{n-k} & = 0, \end{cases}$$

and let the $(n-k) \times n$ matrix $(a_j^i)$ be of maximal rank $n-k < n$. Then, for $x_1$ and $x_2$ running over $L$, the vectors $x_1 - x_2$ form a $k$-dimensional vector space. For this reason, $L$ is called a $k$-dimensional plane in $\mathbb{R}^n$, and a hyperplane for $k = n-1$.

In order to define the length of a straight line segment in $\mathbb{R}^n$, we must introduce a symmetric positive *Euclidean scalar product* on the $n$-dimensional vector space, i.e., a real-valued function $\langle \xi, \eta \rangle$ of a pair of vectors such that:

a) $\langle \lambda_1\xi_1 + \lambda_2\xi_2, \eta \rangle = \lambda_1\langle \xi_1, \eta \rangle + \lambda_2\langle \xi_2, \eta \rangle$, $\langle \xi, \lambda_1\eta_1 + \lambda_2\eta_2 \rangle = \lambda_1\langle \xi, \eta_1 \rangle + \lambda_2\langle \xi, \eta_2 \rangle$ for $\lambda_1, \lambda_2 \in \mathbb{R}$ (bilinearity);

b) $\langle \xi, \eta \rangle = \langle \eta, \xi \rangle$ (symmetry);

c) $\langle \xi, \xi \rangle > 0$ if $\xi \neq 0$ (positivity).

A Cartesian space endowed with a Euclidean scalar product is referred to as a *Euclidean space*. The squared *length* of the vector $\xi$ equals $|\xi|^2 = \langle \xi, \xi \rangle$, and the *angle* $\varphi$ between vectors $\xi$ and $\eta$ is defined by the formula

$$\cos\varphi = \frac{\langle \xi, \eta \rangle}{\sqrt{\langle \xi, \xi \rangle}\sqrt{\langle \eta, \eta \rangle}} = \frac{\langle \xi, \eta \rangle}{|\xi||\eta|}.$$

The validity of this definition, i.e., the inequality

$$-1 \leq \cos\varphi \leq 1,$$

is provided by the well-known inequality

$$\langle \xi, \eta \rangle^2 = \left(\sum \xi_i\eta_i\right)^2 \leq \left(\sum \xi_i^2\right)\left(\sum \eta_i^2\right) = \langle \xi, \xi \rangle\langle \eta, \eta \rangle,$$

which holds since

$$\left(\sum \xi_i^2\right)\left(\sum \eta_i^2\right) - \left(\sum \xi_i\eta_i\right)^2 = \frac{1}{2}\sum_{i,j}(\xi_i\eta_j - \xi_j\eta_i)^2 \geq 0.$$

If $x$ and $y$ are two points in $\mathbb{R}^n$, then the *distance* $\rho(x, y)$ between them is defined as the length of the vector $x - y$:

$$\rho(x, y) \equiv |x - y|.$$

The distance so defined fulfills the following requirements:

1) $\rho(x, y) = \rho(y, x)$;

2) $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ (the *triangle inequality*);

3) $\rho(x, y) \geq 0$ for all $x, y$, and $\rho(x, y) = 0$ if and only if $x = y$.

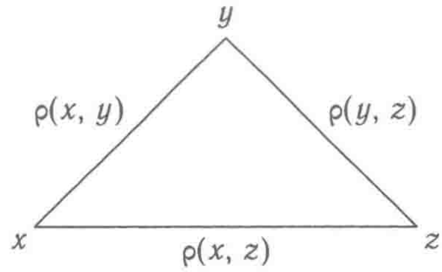


**Figure 1.6.** A triangle.

These definitions of length and angle are based on the Pythagorean theorem. Indeed, by linearity of the scalar product in both arguments,

$$\langle \xi, \eta \rangle = \sum_{1 \le i,j \le n} \xi^i \eta^j \langle e_i, e_j \rangle,$$

where $\xi = \xi^i e_i$, $\eta = \eta^j e_j$. The matrix $(g_{ij})$ with $g_{ij} = \langle e_i, e_j \rangle$ is referred to as the *Gram matrix* of the scalar product. In case this is the identity matrix,

$$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

the coordinates in $\mathbb{R}^n$ are called *Euclidean*, and the basis $e_1, \ldots, e_n$ is said to be orthonormal. When the coordinates are Euclidean, the line segments joining the points $O = (0, \ldots, 0)$, $P = (a, 0, \ldots, 0)$, and $Q = (0, b, 0, \ldots, 0)$ form a right-angle triangle $OPQ$. The lengths of its catheti $OP$ and $OQ$ equal $|a|$ and $|b|$. By the Pythagorean theorem, the length of the side $PQ$ equals $\sqrt{a^2 + b^2}$, which is the length of the difference of radius-vectors of the points $P$ and $Q$. In general, bilinearity and symmetry of the scalar product imply that

$$\langle \xi - \eta, \xi - \eta \rangle = \langle \xi, \xi \rangle + \langle \eta, \eta \rangle - 2\langle \xi, \eta \rangle,$$

and if $\xi$ and $\eta$ are radius-vectors of the points $P$ and $Q$, this identity becomes the cosine theorem for the triangle $OPQ$.

**Theorem 1.2.** *In any finite-dimensional vector space $V$ with a symmetric positive scalar product there exists an orthonormal basis.*

**Proof.** Take a nonzero vector $\xi_1$ and put $e_1 = \xi_1/|\xi_1|$. Denote by $V_1$ the one-dimensional subspace spanned by $\xi_1$, and by $V_1^{\perp}$, its orthogonal complement. The subspace $V_1^{\perp}$ is a hyperplane specified by the linear equation

$$\langle \eta, e_1 \rangle = 0,$$

and its dimension is by one less than that of $V$. Now we perform the same operation with $V_1^{\perp}$: take $\xi_2 \in V_1^{\perp}$, where $\xi_2 \neq 0$, put $e_2 = \xi_2/|\xi_2|$, etc.

Since the space $V$ is finite-dimensional, after finitely many steps we will construct an orthonormal basis $e_1, \dots, e_n$. □

**Corollary 1.1.** *In any Euclidean space there exist Euclidean coordinates.*

In view of this corollary, for any given $n$ we may speak about a single Euclidean space of dimension $n$ because the choice of Euclidean coordinates makes the spaces of the same dimension indistinguishable.

In what follows, when speaking about a Euclidean space, we will assume that the coordinates in this space are Euclidean.

If we are already given some basis $\tilde{e}_1, \dots, \tilde{e}_n$ in the space $V$, then an orthonormal basis can be constructed in a canonical way (by means of the procedure known as the *Gram–Schmidt orthogonalization*). Namely, put $e_1 = \tilde{e}_1/|\tilde{e}_1|$. In the subspace spanned by the vectors $e_1$ and $e_2$, construct an orthogonal basis consisting of $e_1$ and the vector

$$\xi = \tilde{e}_2 - \langle e_1, \tilde{e}_2 \rangle e_1$$

(which is orthogonal to $e_1$ by construction) and then normalize $\xi$ by setting $e_2 = \xi/|\xi|$. Applying this operation successively, for each $k$ we arrive at an orthonormal system of vectors $e_1, \dots, e_k$ such that their linear span coincides with that of the vectors $\tilde{e}_1, \dots, \tilde{e}_k$. Then we put

$$\xi = \tilde{e}_{k+1} - \sum_{i=1}^{k} \langle e_i, \tilde{e}_{k+1} \rangle e_i, \qquad e_{k+1} = \frac{\xi}{|\xi|}.$$

Continuing this way, we end up with an orthonormal basis $e_1, \dots, e_n$ such that for each $k$ the linear span of the vectors $e_1, \dots, e_k$ coincides with the linear span of $\tilde{e}_1, \dots, \tilde{e}_k$.

This canonical procedure can be used, e.g., for constructing bases in the following spaces of polynomials.

Let $\rho$ be a piecewise continuous nonnegative function on an interval $[a, b]$, and let $\rho$ be positive at least on some subinterval of $[a, b]$. Consider the space formed by polynomials $\varphi$, $\psi, \dots$ of degree at most $n$ with scalar product

$$\langle \varphi, \psi \rangle_\rho = \int_a^b \varphi(x)\psi(x)\rho(x)\, dx.$$

The function $\rho$ is called the *weight function*. Two polynomials $\varphi$ and $\psi$ are *orthogonal* if $\langle \varphi, \psi \rangle_\rho = 0$.

Taking the family $\tilde{e}_i = x^{i-1}$, $i = 1, \dots, n+1$, for the initial basis, the canonical orthogonalization procedure results in an orthonormal basis $e_1, \dots, e_{n+1}$ such that $e_i$ for each $i$ is a polynomial of degree $i - 1$. Here are the best-known examples of such systems of orthogonal polynomials.

EXAMPLE 1. Let $[a, b] = [-1, 1]$ and $\rho(x) \equiv 1$. In this case $e_{k+1}(x) = \widetilde{P}_k(x)$, $k \geq 0$. The polynomials

$$P_k(x) = \sqrt{\frac{2}{2k+1}}\, \widetilde{P}_k(x), \qquad \langle \widetilde{P}_j, \widetilde{P}_k \rangle = \delta_{jk},$$

are called the *Legendre polynomials*,

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}\,(3x^2 - 1), \dots,$$
$$P_k(x) = \frac{1}{k!2^k}\,\frac{d^k}{dx^k}\,(x^2 - 1)^k, \dots .$$

EXAMPLE 2. Again, let $[a, b] = [-1, 1]$, but this time take the weight function to be $\rho(x) = 1/\sqrt{1 - x^2}$. Then

$$e_1(x) = \frac{1}{\sqrt{\pi}} T_0(x) = \frac{1}{\sqrt{\pi}}, \qquad e_{k+1}(x) = \sqrt{\frac{2}{\pi}}\, T_k(x), \quad k \geq 1,$$

where $T_k(x) = \cos(k \arccos x)$. The $T_k$ are called the *Chebyshev polynomials.*

If the weight function $\rho(x)$ is defined on the entire real line and decreases at infinity fast enough for the integral

$$\langle \varphi, \psi \rangle_\rho = \int_{-\infty}^{\infty} \varphi(x)\psi(x)\rho(x)\, dx$$

to converge for any pair of polynomials, then this function specifies a scalar product on the space of polynomials on the real line.

EXAMPLE 3. Let $\rho(x) = e^{-x^2/2}$. Then (setting $k! = 1$ for $k = 0$)

$$e_{k+1}(x) = \frac{H_k(x)}{\sqrt{k!}(2\pi)^{1/4}}, \quad k \geq 0,$$

where the polynomials $H_k(x)$ are uniquely determined by the recurrence relations

$$H_0(x) = 1, \qquad H_{k+1}(x) = xH_k(x) - \frac{d}{dx}\, H_k(x).$$

They are called the *Hermite polynomials.*

**1.2.2. The length of a curve.** Now we turn from linear to more complicated objects.

A *smooth parametrized curve* in $\mathbb{R}^n$ is a line in $\mathbb{R}^n$ specified by smooth functions of a univariate parameter $t$:

$$r(t) = (x^1(t), \dots, x^n(t)).$$

For simplicity we restrict ourselves to the cases where the parameter $t$ runs over some interval $[a, b]$ or the entire real line.

The *tangent vector* to (or the *velocity vector* of) the curve at time $t$ is the vector

$$v(t) = \dot{r}(t) = \Big(\frac{dx^1}{dt}, \dots, \frac{dx^n}{dt}\Big).$$

The curve is said to be *regular* if $v(t) \neq 0$ for any $t$.

The *length* of the curve corresponding to variation of the parameter between $a$ and $b$ is defined as the number

$$l = \int_a^b \sqrt{\langle v(t), v(t)\rangle}\, dt = \int_a^b |v(t)|\, dt,$$

i.e., the integral of the length of the velocity vector.

We take this definition of length as an axiom without discussing the justification issues.

If a curve is piecewise smooth, i.e., consists of finitely many smooth curves traced up successively, then its length is defined to be the sum of lengths of these curves.

EXAMPLES. 1. THE STRAIGHT LINE SEGMENT $[P, Q]$. For simplicity, let $a = 0$, $b = 1$, $P = (0, \dots, 0)$, and $\xi = Q = (y^1, \dots, y^n)$, $x^i(t) = y^i t$, $i = 1, \dots, n$. Then $v(t) = (y^1, \dots, y^n)$ and

$$l = \int_0^1 |v(t)|\, dt = \int_0^1 \sqrt{(y^1)^2 + \dots + (y^n)^2}\, dt = \sqrt{(y^1)^2 + \dots + (y^n)^2} = |\xi|.$$

2. THE CIRCLE. Let us specify a circle of radius $R$ by the equations $x^1(t) = R\cos t$, $x^2(t) = R\sin t$, $t \in [0, 2\pi]$. Then $v = (-R\sin t, R\cos t)$ and the length of the circle is equal to

$$l = \int_0^{2\pi} |v|\, dt = \int_0^{2\pi} \sqrt{R^2\sin^2 t + R^2\cos^2 t}\, dt = 2\pi R.$$

3. THE GRAPH OF A FUNCTION. Let a curve on the plane be specified as the graph of a function $x^2 = f(x^1)$ parametrized by the variable $t = x^1$. Then $v = (1, f')$ and

$$l = \int_a^b \sqrt{\langle v, v\rangle}\, dt = \int_a^b \sqrt{1 + f'^2}\, dx^1.$$

Although the formula for the length of a curve involves a parameter, the value of the length does not depend on parametrization.

**Lemma 1.1.** *Let $\tau \in [a', b']$ and let a smooth function $t(\tau)$ be given such that $\frac{dt}{d\tau} > 0$, $\tau(a') = a$, $\tau(b') = b$. Then the curves $r(t) = (x^1(t), \dots, x^n(t))$ and $\tilde{r}(\tau) = (x^1(t(\tau)), \dots, x^n(t(\tau)))$ have the same length.*

**Proof.** Indeed,

$$l(r) = \int_a^b \sqrt{\left\langle \frac{dr}{dt}, \frac{dr}{dt} \right\rangle}\, dt,$$

$$l(\tilde{r}) = \int_{a'}^{b'} \sqrt{\left\langle \frac{d\tilde{r}}{d\tau}, \frac{d\tilde{r}}{d\tau} \right\rangle}\, d\tau = \int_{a'}^{b'} \sqrt{\left\langle \frac{dr}{dt} \cdot \frac{dt}{d\tau}, \frac{dr}{dt} \cdot \frac{dt}{d\tau} \right\rangle}\, d\tau$$

$$= \int_{a'}^{b'} \sqrt{\left\langle \frac{dr}{dt}, \frac{dr}{dt} \right\rangle}\, \frac{dt}{d\tau}\, d\tau = \int_a^b \sqrt{\left\langle \frac{dr}{dt}, \frac{dr}{dt} \right\rangle}\, dt = l(r).$$

The proof is completed. □

## 1.3. Affine transformations

**1.3.1. Matrix formalism. Orientation.** An *affine transformation* of $\mathbb{R}^n$ is an invertible transformation of $\mathbb{R}^n$ into itself which in Cartesian coordinates is given by the formula

$$\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \to A \cdot \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} + \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix} = \begin{pmatrix} a_i^1 x^i + b^1 \\ \vdots \\ a_i^n x^i + b^n \end{pmatrix}, \tag{1.1}$$

where the $n \times n$ matrix $A = (a_j^i)$ and the vector $b = (b^i)$ do not depend on $x^1, \dots, x^n$. In what follows we will briefly write formulas like this as

$$x \to Ax + b \quad \text{or} \quad y = Ax + b.$$

We have already used in 1.1.2 the following rule of summation over repeated indices, which will be used throughout, unless otherwise stated.

*If a formula involves an index repeated once as a superscript and once as a subscript, then this means that the expression is summed up over all possible values of this index.*

For instance, formula (1.1) involves a repeated index $i$, which means summation over all $i = 1, \dots, n$.

Each affine transformation is associated with a linear transformation of the vectors of the Cartesian space, which is specified by the matrix $A$:

$$\xi \to A\xi \quad \text{or} \quad \xi' = A\xi.$$

This transformation has a simple geometric structure. Let the affine transformation map the points $x_1$ and $x_2$ into the points $x_1'$ and $x_2'$. The difference $\xi' = x_1' - x_2'$ depends only on the difference $\xi = x_1 - x_2$ and we have $\xi' = A\xi$.

Any two bases $e_1, \dots, e_n$ and $\tilde{e}_1, \dots, \tilde{e}_n$ of the Cartesian space are related to each other by mutually inverse transition matrices,

$$e_i = a_i^j \tilde{e}_j, \qquad \tilde{e}_j = \tilde{a}_j^i e_i, \qquad a_k^i \tilde{a}_j^k = \delta_j^i. \tag{1.2}$$

The bases are said to have *the same orientation* if $\det A > 0$. Obviously all bases fall into two classes such that:

1) two bases have the same orientation if and only if they belong to the same class;

2) if two bases belong to different classes and $A$ is the transition matrix, then $\det A < 0$.

In order to specify the *orientation* of the space $\mathbb{R}^n$ we have to choose one of these classes as the class of *positively oriented* bases. The bases of the alternative class are then *negatively oriented.*

An affine transformation $x \to Ax + b$ of an oriented Cartesian space is said to be *proper* if the transformation $A\colon (e_1, \dots, e_n) \to (Ae_1, \dots, Ae_n)$ maps positively oriented bases of this space again into positively oriented bases. Obviously, a transformation is proper if and only if $\det A > 0$.

Suppose we are given two systems of Cartesian coordinates $(x^1, \dots, x^n)$ and $(\tilde{x}^1, \dots, \tilde{x}^n)$ in $\mathbb{R}^n$. Assume for simplicity that they have a common origin. Then the coordinates are related by the formula

$$x^i e_i = \tilde{x}^j \tilde{e}_j.$$

Putting (1.2) into this relation we obtain

$$x^i e_i = x^i (a_i^j \tilde{e}_j) = x^i a_i^j \tilde{e}_j = \tilde{x}^j \tilde{e}_j,$$

and therefore

$$\tilde{x}^j = a_i^j x^i.$$

In a similar way it can be shown that

$$x^i = \tilde{a}_j^i \tilde{x}^j. \tag{1.3}$$

In the coordinates $(\tilde{x}^1, \dots, \tilde{x}^n)$ the affine transformation

$$x \to Bx = y$$

is given by another matrix, which can be easily found. The radius-vector of the point $Bx$ is expanded in the basis $e_1, \dots, e_n$ as $Bx = b_k^i x^k e_i$. Therefore, its expansion in the basis $\tilde{e}_1, \dots, \tilde{e}_n$ is

$$y = b_k^i x^k e_i = (b_k^i x^k) a_i^j \tilde{e}_j,$$

and substituting (1.3) we obtain

$$y = (b_k^i \tilde{a}_m^k \tilde{x}^m a_i^j) \tilde{e}_j.$$

Since the matrices $A = (a_j^i)$ and $\widetilde{A} = (\tilde{a}_j^i)$ are mutually inverse, we have

$$\tilde{b}_m^j = a_i^j b_k^i \tilde{a}_m^k, \qquad \widetilde{B} = ABA^{-1}.$$

Now the formula for the transformation $x \to Bx$ in the new coordinates takes its final form

$$\tilde{x} \to \widetilde{B}\tilde{x}, \qquad \widetilde{B} = ABA^{-1}.$$

Usually we will tacitly assume that a coordinate system is given. In this case we will identify transformations with matrices which specify them. These are always nonsingular square matrices.

If we dispense with these properties and consider arbitrary matrices, we obtain linear mappings from one space into another, with possibly different dimensions of the spaces.

A mapping $A\colon V \to W$ of vector spaces is said to be *linear* if

1) $A(\xi_1 + \xi_2) = A\xi_1 + A\xi_2$ for any pair of vectors $\xi_1, \xi_2 \in V$;

2) $A(\lambda\xi) = \lambda A\xi$ for any $\xi \in V$ and $\lambda \in \mathbb{R}$.

If $e_1, \dots, e_n$ is a basis in $V$ and $\tilde{e}_1, \dots, \tilde{e}_m$ a basis in $W$, then a linear mapping is uniquely determined by the matrix $a^i_j$ such that

$$A\xi = a^i_j \xi^j \tilde{e}_i, \quad \text{where } \xi = \xi^j e_j.$$

If $V = W$ and the mapping $A$ is invertible, we obtain a *linear transformation.*

**1.3.2. Affine group.** Recall that a *group* is a set $G$ endowed with two operations on its elements: multiplication associating with each ordered pair of elements $g, h$ their product $gh \in G$, and inversion associating with each element $g$ its inverse $g^{-1} \in G$.

Moreover, there is the unit element $1 \in G$ (the identity of $G$), and all the elements $f, g, h \in G$ satisfy the following conditions:

1) $f(gh) = (fg)h$ (associativity);

2) $gg^{-1} = g^{-1}g = 1$;

3) $g \cdot 1 = 1 \cdot g = g$.

The *composition* of mappings $\varphi\colon X \to Y$ and $\psi\colon Y \to Z$ is the mapping

$$\psi\varphi\colon X \to Z$$

such that

$$\psi\varphi(x) = \psi(\varphi(x))$$

for every $x \in X$. The composition operation is associative,

$$\chi(\psi\varphi) = (\chi\psi)\varphi.$$

The composition of affine transformations $\varphi\colon x \to Ax+c$ and $\psi\colon x \to Bx+d$ is

$$\psi\varphi\colon x \to BAx + (Bc + d),$$

or, in more detail,

$$\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \to \begin{pmatrix} b^1_i a^i_j x^j + b^1_i c^i + d^1 \\ \vdots \\ b^n_i a^i_j x^j + b^n_i c^i + d^n \end{pmatrix}.$$

The composition of affine transformations is obviously invertible and

$$(\psi\varphi)^{-1} = \varphi^{-1}\psi^{-1}.$$

Therefore, the composition of affine transformations is again an affine transformation.

**Theorem 1.3.** *Affine transformations with composition and inversion operations form a group with the identity transformation as the unit element.*

The group of affine transformations is called the *affine group* and is denoted by $\mathrm{A}(n)$.

**Proof of the theorem.** Since the affine transformation $g\colon x \to Ax + b$ is invertible, we have $Ax+b \neq Ay+b$ for $x \neq y$, hence $A(x-y) \neq 0$. Therefore, the matrix $A$ is invertible, so the transformation

$$g^{-1}\colon x \to A^{-1}x - A^{-1}b$$

is also affine. The proof is completed. □

Affine transformations of the form

$$x \to x + b$$

are called *translations.* A translation $x \to x + a$ is a shift along the vector $a$. The product of translations along vectors $a$ and $b$ is the translation along the vector $a + b$, and the inverse of the translation along the vector $a$ is the translation by $(-a)$. Hence we conclude that translations form a group, which is isomorphic to the additive group of vectors of the space. This group is *commutative*, i.e., the order of factors does not affect the value of the product, $gh = hg$.

Another example of affine transformations is provided by *linear* transformations, which are written in Cartesian coordinates as

$$x \to Ax.$$

The composition of linear transformations $x \to Ax$ and $x \to Bx$ is again a linear transformation

$$x \to BAx.$$

The inverse transformation to $x \to Ax = y$ is $y \to A^{-1}y = x$.

The linear transformations of $\mathbb{R}^n$ form a (*complete*) *linear group*, which coincides with the group of nonsingular $n \times n$ matrices denoted by $\mathrm{GL}(n)$. For $n > 1$ this group is noncommutative.

A special class of linear transformations is formed by *dilatations* or *homotheties* determined by the matrices $A$ that are multiples of the identity matrix,

$$x \to \lambda x = y.$$

A *motion* of the space $\mathbb{R}^n$ is a smooth mapping $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ preserving distances between points,

$$|x - y| = \rho(x, y) = \rho(\varphi(x), \varphi(y)) = |\varphi(x) - \varphi(y)|$$

for any pair of points $x$ and $y$.

It is easily seen that translations preserve distances between points, so that they are a particular case of motions.

**Theorem 1.4.** *Any motion of the Euclidean space is an affine transformation.*

**Proof.** It can be easily shown that any motion $\varphi$ maps straight lines into straight lines. We will present this geometric argument, although it will not be needed in the proof.

Let $x$, $y$, $z$ be three points lying consecutively on a straight line. Construct the triangle with vertices at $\varphi(x)$, $\varphi(y)$, $\varphi(z)$. By the triangle inequality,

$$\rho(\varphi(x), \varphi(y)) + \rho(\varphi(y), \varphi(z)) \geq \rho(\varphi(x), \varphi(z)). \tag{1.4}$$

The equality is attained only when the angle at the vertex $\varphi(y)$ equals $\pi$, i.e., all the three vertices lie on the same straight line (see Figure 1.7).

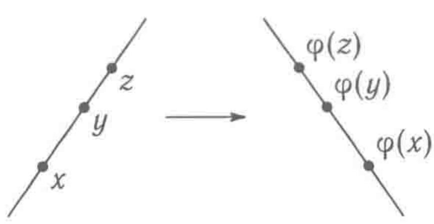


**Figure 1.7.** The extreme case in inequality (1.4).

Since

$$\rho(x, y) + \rho(y, z) = \rho(x, z)$$

and $\varphi$ is a motion, we see that (1.4) is also an equality. The point $y$ inside the interval $xz$ was chosen arbitrarily. Therefore, $\varphi$ maps the entire straight line segment again into a straight line segment, hence it maps straight lines into straight lines.

Let $\varphi(0) = b$, where $0 = (0, \dots, 0)$. The composition of the transformation $\varphi$ and the translation $x \to x - b$ is a motion

$$\psi\colon x \to \varphi(x) - b$$

such that $\psi(0) = 0$. We will prove that the motion $\psi$ is linear.

First of all, observe that if a motion $\alpha$ maps the zero point into itself, then

$$\langle x, y\rangle = \langle \alpha(x), \alpha(y)\rangle$$

for any vectors $x$ and $y$. Indeed, the condition that $\alpha$ preserves the distance between $x$ and $y$ can be written as

$$\langle x - y, x - y\rangle = \langle \alpha(x) - \alpha(y), \alpha(x) - \alpha(y)\rangle,$$

and for $y = 0$ the equality $\alpha(0) = 0$ implies

$$\langle x, x\rangle = \langle \alpha(x), \alpha(x)\rangle.$$

Generally, we have

$$\begin{aligned}\langle x - y, x - y\rangle &= \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle \\ &= \langle \alpha(x), \alpha(x)\rangle - 2\langle \alpha(x), \alpha(y)\rangle + \langle \alpha(y), \alpha(y)\rangle \\ &= \langle \alpha(x) - \alpha(y), \alpha(x) - \alpha(y)\rangle,\end{aligned}$$

which implies that $\langle x, y\rangle = \langle \alpha(x), \alpha(y)\rangle$.

Now, let $x_1, \dots, x_n$ be the points with radius-vectors $e_1 = (1, 0, \dots, 0)$, $\dots$, $e_n = (0, \dots, 0, 1)$. Denote by $\tilde{e}_i$ the radius-vector of the point $\psi(x_i)$. Since $\psi$ is a motion and $\psi(0) = 0$, we have

$$\begin{gathered}\langle e_i, e_i\rangle = \langle \tilde{e}_i, \tilde{e}_i\rangle = 1, \\ \langle \tilde{e}_i, \tilde{e}_j\rangle = \langle e_i, e_j\rangle = 0 \quad \text{for } i \neq j.\end{gathered}$$

Therefore, the vectors $\tilde{e}_1, \dots, \tilde{e}_n$ form an orthonormal basis.

Denote by $A$ the matrix specifying the linear transformation which maps $\tilde{e}_i$ into $e_i$. Since this transformation maps an orthonormal basis again into an orthonormal basis, it determines a motion $\chi\colon x \to Ax$. Obviously, the composition $\chi\psi$ is also a motion, and

$$\chi\psi(0) = 0, \qquad \chi\psi(e_i) = e_i.$$

We expand the vectors $x$ and $\chi\psi(x)$ in the basis $e_1, \dots, e_n$:

$$x = \sum_{i=1}^{n} \langle x, e_i\rangle e_i, \qquad \chi\psi(x) = \sum_{i=1}^{n} \langle \chi\psi(x), e_i\rangle e_i.$$

Since

$$\langle \chi\psi(x), e_i\rangle = \langle \chi\psi(x), \chi\psi(e_i)\rangle = \langle x, e_i\rangle,$$

we have

$$x = \chi\psi(x)$$

for any vector $x$. Hence the motion $\chi\psi$ is the identity, and therefore the transformation $\psi$ is linear,

$$\psi\colon x \to A^{-1}x.$$

This implies that the transformation $\varphi$ is linear as well,

$$\varphi(x) = A^{-1}x + b.$$

Hence the theorem. □

It is easily seen that the motions of the space $\mathbb{R}^n$ form a group. This group is denoted by $\mathrm{E}(n)$.

Recall that a *subgroup* $F$ of a group $G$ is a set of elements of $G$ which is closed with respect to the multiplication and inversion operations. In this case $F$ is itself a group relative to these operations. The groups of translations, linear transformations, and motions are subgroups of the affine group $\mathrm{A}(n)$ and are denoted by $\mathbb{R}^n$, $\mathrm{GL}(n)$, and $\mathrm{E}(n)$, respectively.

A mapping $\varphi$ of a group $G$ into a group $H$,

$$\varphi\colon G \to H,$$

is a *homomorphism* of groups if it preserves the operations: $\varphi(ab) = \varphi(a)\varphi(b)$ and $\varphi(a^{-1}) = (\varphi(a))^{-1}$. Obviously, it must map the unit element into the unit element: $\varphi(1_G) = \varphi(a)\varphi(a^{-1}) = 1_H$.

The elements of $G$ that are mapped into $1_H$ by the homomorphism $\varphi$ form a subgroup $\operatorname{Ker}\varphi$, which is called the *kernel* of the homomorphism.

If, for a subgroup $F \subset G$, there is a homomorphism $\varphi\colon G \to H$ into a group $H$ such that $\operatorname{Ker}\varphi = F$, then the subgroup $F$ is called a *normal subgroup* (or a *normal divisor*). In such a case, if the homomorphism $\varphi$ maps the group $G$ onto the entire group $H$, then $H$ is called the *quotient group* of the group $G$ by the subgroup $F$, $H = G/F$. Obviously, to any normal subgroup $F$ corresponds the quotient group $G/F = \varphi(G)$.

If for a homomorphism $\varphi\colon G \to H$ there is a homomorphism $\varphi^{-1}\colon H \to G$ such that the mappings $\varphi^{-1}\varphi\colon G \to G$ and $\varphi\varphi^{-1}\colon H \to H$ are the identities, then the homomorphism $\varphi$ is called an *isomorphism*, and the groups $G$ and $H$ are said to be *isomorphic*.

The simplest normal subgroup of the group $\mathrm{GL}(n)$ is the *special linear group* $\mathrm{SL}(n)$ consisting of $n \times n$ matrices with determinant 1. The subgroup $\mathrm{SL}(n)$ is the kernel of the homomorphism which associates with each matrix in $\mathrm{GL}(n)$ its determinant,

$$\det\colon \mathrm{GL}(n) \to \mathbb{R}^*,$$

where $\mathbb{R}^*$ denotes the multiplicative group of nonzero real numbers. The groups $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$ are also denoted by $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{SL}(n, \mathbb{R})$, respectively.

**Theorem 1.5.** *The mapping that associates with each affine transformation*

$$x \to Ax + b$$

*the linear transformation*

$$x \to Ax$$

*is a homomorphism of the affine group $G = \mathrm{A}(n)$ onto the group of linear transformations $H = \mathrm{GL}(n)$. The kernel of this homomorphism coincides with the group of translations of the space $\mathbb{R}^n$.*

**Proof.** The composition of transformations $x \to A_1x + b_1$ and $x \to A_2x + b_2$ is $x \to A_2A_1x + (A_2b_1 + b_2)$, hence it is mapped into the linear transformation $x \to A_2A_1x$. The inverse transformation to $x \to Ax + b$ is

$$x \to A^{-1}x - A^{-1}b,$$

which is mapped into the linear transformation

$$x \to A^{-1}x.$$

Therefore, this mapping of the affine group into the group of linear transformations is a homomorphism of groups. It is obvious that the kernel of this homomorphism consists precisely of translations. □

**Corollary 1.2.** *The group of translations of the space $\mathbb{R}^n$ is a normal divisor of the affine group $\mathrm{A}(n)$. Hence it is a normal divisor of the group of motions $\mathrm{E}(n)$.*

**1.3.3. Motions of Euclidean spaces.** By Theorem 1.4, any motion of the affine space $\mathbb{R}^n$ is an affine transformation. Consider the restriction of the homomorphism in Theorem 1.5 to the group of motions. The image of the group of motions under this homomorphism is the group of *orthogonal transformations.*

The orthogonality conditions can be written down algebraically. Choose some Euclidean coordinates in the space. Then the transformation $x \to Ax + b$ is a motion of the space $\mathbb{R}^n$ if and only if

$$\langle A\xi, A\xi \rangle = \langle \xi, \xi \rangle$$

for each vector $\xi$. Indeed, the distance between the points $x_1$ and $x_2$ equals the length of the vector $\xi = x_1 - x_2$, while the distance between the images of the points $x_1$ and $x_2$ equals the length of the vector $A\xi$. Therefore, the

distance between $x_1$ and $x_2$ is preserved under an affine transformation if and only if $|A\xi| = |\xi|$, or, equivalently,

$$|A\xi|^2 = \langle A\xi, A\xi\rangle = \langle \xi, \xi\rangle = |\xi|^2.$$

We can write down this equality in the matrix form as

$$\xi^\top A^\top A\xi = \xi^\top \xi,$$

where $\xi$ is written as a column vector and $\top$ means transposition (in particular, $(a^\top)^i_j = a^j_i$). Since this equality is fulfilled for all $\xi$, we obtain

$$A^\top A = 1, \tag{1.5}$$

where 1 is the $n \times n$ identity matrix.

A square matrix $A$ satisfying equation (1.5) is said to be *orthogonal.* Orthogonal $n \times n$ matrices form the *orthogonal group* $\mathrm{O}(n)$. Since $\det A^\top = \det A$, equality (1.5) implies that $\det A = \pm 1$ for $A \in \mathrm{O}(n)$. The matrices in $\mathrm{O}(n)$ with determinant equal to 1 form the subgroup of orthogonal transformations preserving orientation. This subgroup is denoted by $\mathrm{SO}(n)$.

Thus we have proved the following assertion.

**Lemma 1.2.** *An affine transformation written in Euclidean coordinates as*

$$x \to Ax + b$$

*is a motion if and only if the matrix $A$ is orthogonal.*

Suppose we are given a family of orthogonal matrices $B(t)$, which are smooth functions of a parameter $t$.

**Theorem 1.6.** *We have*

$$\frac{dB}{dt}(t) = A(t)B(t),$$

*where $A(t)$ is a skew-symmetric matrix depending on $t$.*

**Proof.** Let $B(0) = B_0$. We write down the Taylor expansion of $B(t)$ about the point $t = 0$:

$$B(t) = (1 + At + O(t^2))B_0.$$

Substituting this expansion into (1.5) we obtain

$$B_0^\top (A^\top + A)B_0 t + O(t^2) = 0,$$

hence $A^\top + A = 0$, so that the matrix $A = \frac{dB}{dt} B^{-1}$ is skew-symmetric. □

Consider the motions of two- and three-dimensional Euclidean spaces in more detail. We assume that our coordinates are Euclidean.

Let $n = 2$. The orthogonality condition for a matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ becomes

$$a^2 + c^2 = b^2 + d^2 = 1, \qquad ab + cd = 0.$$

The solutions to this system fall into one of two classes:

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}.$$

The first class consists of rotations (through an angle $\varphi$) about the origin.

**Lemma 1.3.** *A linear transformation $x \to Ax$ with*

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

*is a reflection in a straight line. In suitably chosen Euclidean coordinates this transformation can be written as*

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}.$$

**Proof.** For $\cos\varphi = 1$ this transformation is the reflection in the line $x^2 = 0$, and for $\cos\varphi = -1$ this is the reflection in the line $x^1 = 0$.

Let $\cos\varphi \neq \pm 1$. The roots of the equation $\det(A - \lambda \cdot 1) = 0$ are $\lambda = \pm 1$, and one can easily construct a basis of mutually orthogonal eigenvectors:

$$\tilde{e}_1 = (\sin\varphi, 1 - \cos\varphi)^\top, \qquad \tilde{e}_2 = (\sin\varphi, -(1 + \cos\varphi))^\top,$$
$$A\tilde{e}_1 = \tilde{e}_1, \qquad A\tilde{e}_2 = -\tilde{e}_2.$$

Therefore, $A$ is the reflection in the line $\{t\tilde{e}_1 : t \in \mathbb{R}\}$. □

Thus, if an orthogonal transformation of the plane is proper, i.e., $\det A = 1$, then it is a rotation; otherwise it is the reflection in a straight line.

The following theorem provides a classification of motions of the plane.

**Theorem 1.7.** 1. *Any proper motion of the plane is either a rotation about a fixed point or a translation.*

2. *Any nonproper motion of the plane $\mathbb{R}^2$ is a "glide-reflection", i.e., a composition of the reflection in a straight line and a translation along this line.*

**Proof.** Let $\det A = 1$. If $A$ is the identity matrix, $A = 1$, then the motion is a translation. Otherwise, if

$$A = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}, \quad \cos\varphi \neq 1,$$

the equation

$$Ax + b = x$$

has a solution because the matrix $(1 - A)$ is invertible, and the solution $x_0$ has the form $x_0 = (1 - A)^{-1}b$. Shifting the origin into the point $x_0$ we

obtain that in the new coordinates the motion is specified as $\tilde{x} \to A\tilde{x}$, and therefore, it is a rotation about the point $x_0$.

Now let $\det A = -1$. Using Lemma 1.3 we represent the motion as

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} b^1 \\ b^2 \end{pmatrix}.$$

Now we shift the origin into the point $\left(0, \frac{b^2}{2}\right)$. Then in the new coordinates $\tilde{x}^1 = x^1$, $\tilde{x}^2 = x^2 - \frac{b^2}{2}$ the motion takes the form

$$\begin{pmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{pmatrix} \to \begin{pmatrix} \tilde{x}^1 + b^1 \\ -\tilde{x}^2 \end{pmatrix},$$

which shows that it is a glide-reflection. The proof is completed. □

Now let $n = 3$.

**Lemma 1.4.** *Any linear transformation of the space $\mathbb{R}^3$ has a nonzero eigenvector, i.e., a vector $\xi$ such that*

$$A\xi = \lambda\xi \quad \textit{and} \quad \xi \neq 0.$$

**Proof.** Any eigenvector $\xi$ is a solution to the equation $(A - \lambda \cdot 1)\xi = 0$. This equation has nonzero solutions if and only if $\det(A - \lambda \cdot 1) = 0$. Since $p(\lambda) = \det(A - \lambda \cdot 1)$ is a polynomial of odd degree, it has a real root $\lambda_0$. Now take for $\xi$ a nonzero solution to the equation

$$(A - \lambda_0 \cdot 1)\xi = 0.$$

The proof is completed. □

Let $\xi$ be an eigenvector of an orthogonal transformation $x \to Ax$. Then $\lambda = \pm 1$ because the transformation preserves the length of the vector $\xi$, i.e., $|A\xi| = |\xi|$.

If $\langle \xi, \eta \rangle = 0$, then $\langle A\xi, A\eta \rangle = \langle \xi, \eta \rangle = 0$ and $\langle \xi, A\eta \rangle = 0$ because $\langle A\xi, A\eta \rangle = \pm\langle \xi, A\eta \rangle$. Therefore, the transformation $A$ maps the plane $\langle \xi, \eta \rangle = 0$ orthogonal to the vector $\xi$ into itself.

Take a new orthonormal basis such that $e_1 = \xi/|\xi|$ and the vectors $e_2$ and $e_3$ are orthogonal to $\xi$. In appropriate Euclidean coordinates this transformation has the form

$$x \to \begin{pmatrix} \pm 1 & 0 \\ 0 & A' \end{pmatrix} x,$$

where the plane transformation $x' \to A'x'$ is also orthogonal.

It follows from the classification of orthogonal transformations of a plane that $A'$ is either a rotation about the origin or a reflection in a straight line.

Hence we obtain two statements.

1. If $\det A = 1$, then the matrix $A$ can be reduced to the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

In the former case this is the rotation through the angle $\varphi$ about the $x^1$-axis, and in the latter case the rotation through the angle $\pi$ about the $x^2$-axis.

2. If $\det A = -1$, then the matrix $A$ can be reduced to the form

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

In the former case this is a *rotatory reflection*, i.e., the composition of a rotation (through the angle $\varphi$) about some axis (presently the $x^1$-axis) and the reflection in a plane orthogonal to this axis. The latter case is also a rotatory reflection, namely, the one about the $x^3$-axis through the angle $\varphi = 0$.

Thus we have proved the following assertion.

**Lemma 1.5.** 1. *If an orthogonal transformation $x \to Ax$ of $\mathbb{R}^3$ is proper (i.e., $\det A = 1$), then it is a rotation about some axis through some angle.*

2. *If an orthogonal transformation of $\mathbb{R}^3$ is not proper (i.e., $\det A = -1$), then it is a rotatory reflection.*

Now we turn to general (not necessarily origin-preserving) motions of $\mathbb{R}^3$. The preceding lemma implies that a proper motion reduces to the form

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} + \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix}.$$

From the classification of motions in the plane we see that if $\cos\varphi = 1$, then the motion is a translation; otherwise, if $\cos\varphi \neq 1$, then the mapping

$$\begin{pmatrix} x^2 \\ x^3 \end{pmatrix} \to \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x^2 \\ x^3 \end{pmatrix} + \begin{pmatrix} b^2 \\ b^3 \end{pmatrix}$$

has a fixed point in the plane $x^1 = 0$. Shifting the origin to this point reduces the motion of $\mathbb{R}^3$ to the form

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} + \begin{pmatrix} b^1 \\ 0 \\ 0 \end{pmatrix}.$$

Such motions are referred to as *screw-displacements*; they include, as particular cases, translations (for $\varphi = 0$) and rotations about an axis (for $b = 0$).

Now we apply similar reasoning to nonproper motions, which in general reduce to the form

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \to \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} + \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix}.$$

For $\cos\varphi = 1$ this is a *glide-reflection*, i.e., the composition of a reflection in the plane $x^1 = b^1/2$ and a translation in the direction parallel to this plane. For $\cos\varphi \neq 1$ we can reduce this motion (possibly, shifting the origin) to the form

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \to \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} + \begin{pmatrix} b^1 \\ 0 \\ 0 \end{pmatrix}.$$

Shifting the origin into the point $(b^1/2, 0, 0)$, we obtain that any nonproper motion of $\mathbb{R}^3$ can be reduced to the form

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} \to \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix},$$

i.e., such a motion is a rotatory reflection.

Thus we have proved the following theorem.

**Theorem 1.8.** *Any proper motion of the space $\mathbb{R}^3$ is a screw-displacement, and any nonproper motion is either a rotatory reflection or a glide-reflection.*

## 1.4. Curves in Euclidean space

**1.4.1. The natural parameter and curvature.** We have defined smooth parametrized curves in $\mathbb{R}^n$ as smooth mappings of intervals or the entire real line into $\mathbb{R}^n$. In what follows, we treat parametrized curves obtainable by tracing out the same line with different velocities as one and the same curve.

Namely, let $r_1(t)$ and $r_2(\tau)$ be smooth parametrized curves, and let there exist a function $\tau(t)$ such that $r_1(t) = r_2(\tau(t))$ for all $t$. We assume that $d\tau/dt > 0$ and $\tau(t)$ takes all possible values of $\tau$ when $t$ varies. In this case we will treat $r_1(t)$ and $r_2(\tau)$ as the same curve with different parameters $t$ and $\tau$.

A parameter $l$ of a curve $r(l)$ is said to be *natural* if the length of the arc of the curve corresponding to variation of $l$ from $l'$ to $l''$ equals $l'' - l'$ for any pair $l'$ and $l''$:

$$\int_{l'}^{l''} \left| \frac{dr}{dl} \right| dl = l'' - l'.$$

This formula implies that $l$ is natural if and only if

$$\left| \frac{dr}{dl} \right| \equiv 1.$$

**Lemma 1.6.** *For every regular curve there exists a natural parameter.*

**Proof.** Let $r(t)$ be a parametrization of the curve such that $dr/dt \neq 0$ everywhere. A parameter $l = l(t)$ is natural if only if

$$\left|\frac{dr}{dl}\right| = \left|\frac{dr}{dt}\right| \cdot \frac{dt}{dl} = \left|\frac{dr}{dt}\right| \Big/ \frac{dl}{dt} = 1.$$

Hence we define $l = l(t)$ as a solution to the differential equation

$$\frac{dl}{dt} = \left|\frac{dr}{dt}\right|.$$

Since the right-hand side of this equation is a smooth function, it has a solution, which is unique up to an additive constant, with properties $dl/dt > 0$ and $|dr/dl| \equiv 1$. Hence the lemma. $\square$

The following lemma describes an important property of the natural parameter.

**Lemma 1.7.** *The acceleration vector $d^2r/dl^2$ of a naturally parametrized curve is everywhere orthogonal to the velocity vector $v = dr/dl$:*

$$\frac{d^2r}{dl^2} \perp \frac{dr}{dl}.$$

**Proof.** The proof consists in a simple calculation:

$$\frac{d}{dl}\left\langle \frac{dr}{dl}, \frac{dr}{dl} \right\rangle = 0,$$

since $\left|\frac{dr}{dl}\right| \equiv 1$. Therefore

$$\frac{d}{dl}\left\langle \frac{dr}{dl}, \frac{dr}{dl} \right\rangle = 2\left\langle \frac{d^2r}{dl^2}, \frac{dr}{dl} \right\rangle = 0,$$

which was to be shown. $\square$

The length of the acceleration vector (relative to the natural parameter on the curve) is called the *curvature* of the curve:

$$k = \left|\frac{d^2r}{dl^2}\right|,$$

and the reciprocal quantity is the *radius of curvature*:

$$R = \frac{1}{k}.$$

Now we consider curves in $\mathbb{R}^2$ and $\mathbb{R}^3$ in more detail.

**1.4.2. Curves on the plane.** Select some orientation of the plane $\mathbb{R}^2$. Let $r(l)$ be a naturally parametrized curve. With each point of this curve we associate the orthonormal basis $(v, n)$ by the following rule:

1) $v = dr/dl$,

2) the basis $(v, n)$ is positively oriented.

This basis in $\mathbb{R}^2$ is called the *Frenet basis* (*frame*). The vector $n$ is the *normal* to the curve.

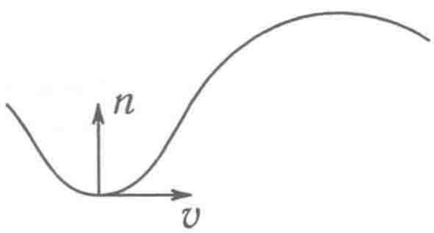


**Figure 1.8.** Frenet basis on the plane.

Lemma 1.7 implies that

$$\frac{d^2r}{dl^2} = kn,$$

where $|k|$ is the curvature defined above. In this way we define the *signed curvature* $k$ of a planar curve.

EXAMPLE. Let $(x(l) = R\cos(l/R),\ y(l) = R\sin(l/R))$ be the circle of radius $R$ with the natural parameter. Then

$$\frac{d^2r}{dl^2} = \Big(-\frac{1}{R}\cos\frac{l}{R}, -\frac{1}{R}\sin\frac{l}{R}\Big);$$

hence the curvature of a circle is constant and equal to $1/R$, where $R$ is its radius.

For a straight line, the velocity vector is constant, hence the line has zero curvature everywhere.

The deformations of Frenet frames are described by the *Frenet equations.*

**Theorem 1.9.** *For a smooth regular planar curve $r(l)$ the following Frenet equations hold:*

$$\frac{d}{dl}\begin{pmatrix} v \\ n \end{pmatrix} = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix}\begin{pmatrix} v \\ n \end{pmatrix}.$$

**Proof.** Since $\langle v, v\rangle = \langle n, n\rangle = 1$, $\langle v, n\rangle = 0$, we have the following relations:

$$\frac{d}{dl}\langle v, v\rangle = 2\Big\langle \frac{dv}{dl}, v\Big\rangle = 0, \qquad \frac{d}{dl}\langle n, n\rangle = 2\Big\langle \frac{dn}{dl}, n\Big\rangle = 0,$$

$$\frac{d}{dl}\langle v, n\rangle = \Big\langle \frac{dv}{dl}, n\Big\rangle + \Big\langle v, \frac{dn}{dl}\Big\rangle = 0.$$

But the vectors $v$ and $n$ form an orthonormal basis. Hence the first two equalities imply

$$\frac{dv}{dl} = \alpha n, \qquad \frac{dn}{dl} = \beta v,$$

while the third yields

$$\langle \alpha n, n\rangle + \langle v, \beta v\rangle = \alpha + \beta = 0.$$

By definition, $\alpha = k$, hence $\beta = -k$. The proof is completed. □

There is a one-to-one correspondence between smooth curvatures $k(l)$ and curves considered up to proper motions.

**Theorem 1.10.** 1. *For any sufficiently smooth function* $k\colon [0, l] \to \mathbb{R}$ *there exists a smooth naturally parametrized curve* $r\colon [0, l] \to \mathbb{R}^2$ *with curvature equal to* $k(l)$. *This curve can be found from the Frenet equations.*

2. *If the curvatures of naturally parametrized planar curves* $r_1(l)$ *and* $r_2(l)$ *are equal, then there exists a proper motion* $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ *which maps one curve into the other,* $r_2(l) = \varphi(r_1(l))$.

**Proof.** 1. Take an orthonormal positively oriented basis $e_1, e_2$ and consider a solution to the Frenet equations with initial conditions $v(0) = e_1$, $n(0) = e_2$. Such a solution exists and is unique because the Frenet equations may be viewed as an ordinary differential equation in $\mathbb{R}^4 = \mathbb{R}^2 \oplus \mathbb{R}^2$.

At the same time the scalar products of $v(l)$ and $n(l)$ satisfy another system of differential equations,

$$\frac{d\langle v, v\rangle}{dl} = 2k\langle v, n\rangle, \quad \frac{d\langle n, n\rangle}{dl} = -2k\langle v, n\rangle, \quad \frac{d\langle v, n\rangle}{dl} = k\langle n, n\rangle - k\langle v, v\rangle.$$

This system with initial conditions $\langle v, v\rangle = \langle n, n\rangle = 1$, $\langle v, n\rangle = 0$ has a unique solution. It is easily seen that this solution is a constant, hence for any $l$ the vectors $v(l)$ and $n(l)$ form an orthonormal basis in $\mathbb{R}^2$.

Define the curve by the formula

$$r(l) = \int_0^l v(t)\, dt.$$

By construction, $|v(l)| = 1$; therefore, $l$ is a natural parameter, $v$ is the velocity vector, and $(v, n)$ is the Frenet frame.

Since $dv/dl = kn$, we see that $k(l)$ is the curvature of the planar curve $r(l)$.

2. Let $(v_i, n_i)$ be the Frenet frame of the curve $r_i$, $i = 1, 2$. We may assume for definiteness that $0 \le l \le L$. Define $\varphi$ as the composition of the shift by $a = r_2(0) - r_1(0)$ and the linear transformation $A$ that maps $(v_1(0), n_1(0))$ into $(v_2(0), n_2(0))$,

$$\varphi(x) = A(x) + a.$$

The Frenet frame of the curve $\varphi(r_1)$ is $(Av_1, An_1)$. It satisfies the equation

$$\frac{d}{dl}\begin{pmatrix} Av_1 \\ An_1 \end{pmatrix} = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix}\begin{pmatrix} Av_1 \\ An_1 \end{pmatrix},$$

and so does the frame $(v_2, n_2)$. Moreover, these two frames coincide for $l = 0$. Since the Frenet equations with given initial conditions have a unique solution, the equalities $Av_1 = v_2$, $An_1 = n_2$ hold everywhere.

By the choice of the shift we have

$$r_2(l) = r_2(0) + \int_0^l v_2(t)\,dt = \varphi(r_1(l)).$$

The proof is completed. □

For planar curves the Frenet equations can be easily integrated. Let $\alpha(l) = \int_0^l k(t)\,dt$. Then the curvature of the curve

$$r(l) = \left(\int_0^l \cos\alpha(t)\,dt,\ \int_0^l \sin\alpha(t)\,dt\right)$$

is $k(l)$. For curves in $\mathbb{R}^3$ the Frenet equations cannot be integrated explicitly.

**1.4.3. Curvature and torsion of curves in $\mathbb{R}^3$.** In the oriented Euclidean space $\mathbb{R}^3$ there is one more binary operation besides the scalar product, namely, the *vector* (or *cross*) product. In Euclidean coordinates it has the form

$$[\xi, \eta] = \det\begin{pmatrix} e_1 & e_2 & e_3 \\ \xi^1 & \xi^2 & \xi^3 \\ \eta^1 & \eta^2 & \eta^3 \end{pmatrix}, \quad \text{i.e., } [e_1, e_2] = e_3,\ [e_2, e_3] = e_1,\ [e_3, e_1] = e_2,$$

where $(e_1, e_2, e_3)$ is a positively oriented orthonormal basis, $\xi = \xi^i e_i$, $\eta = \eta^j e_j$, and the determinant is calculated by the formal algebraic rules:

$$[\xi, \eta] = (\xi^2\eta^3 - \xi^3\eta^2)e_1 + (\xi^3\eta^1 - \xi^1\eta^3)e_2 + (\xi^1\eta^2 - \xi^2\eta^1)e_3.$$

This product is linear in both factors,

$$[\lambda_1\xi_1 + \lambda_2\xi_2, \eta] = \lambda_1[\xi_1, \eta] + \lambda_2[\xi_2, \eta],$$
$$[\xi, \lambda_1\eta_1 + \lambda_2\eta_2] = \lambda_1[\xi, \eta_1] + \lambda_2[\xi, \eta_2]$$

for $\lambda_1, \lambda_2 \in \mathbb{R}$, and anticommutative,

$$[\xi, \eta] = -[\eta, \xi].$$

Moreover, it can be easily checked that

$$\langle [\xi,\eta],[\xi,\eta]\rangle = \langle \xi,\xi\rangle\langle \eta,\eta\rangle - \langle \xi,\eta\rangle^2,$$
$$\langle [\xi,\eta],\xi\rangle = \langle [\xi,\eta],\eta\rangle = 0.$$

These equations determine $[\xi,\eta]$ up to a sign. Indeed, if the vectors $\xi$ and $\eta$ are proportional to each other, then $[\xi,\eta]=0$ by anticommutativity. Otherwise $|[\xi,\eta]|\neq 0$, the vector $[\xi,\eta]$ is orthogonal to $\xi$ and $\eta$, and we know its length. The sign is chosen so that $(\xi,\eta,[\xi,\eta])$ be a positively oriented basis in $\mathbb{R}^3$.

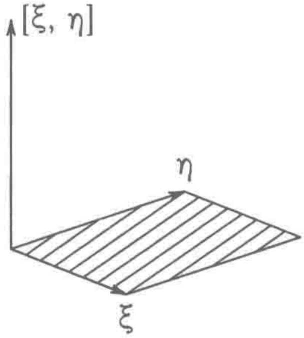


**Figure 1.9.** Vector product.

In addition, the vector product satisfies another important identity, called the *Jacobi identity*,

$$[\xi,[\eta,\chi]] + [\eta,[\chi,\xi]] + [\chi,[\xi,\eta]] = 0$$

for any $\xi$, $\eta$, $\chi$. To prove this identity, it suffices, using linearity and anticommutativity, to reduce it to the case $\xi = e_1$, $\eta = e_2$, $\chi = e_3$. But then it is obvious because

$$[e_1,e_2] = e_3, \qquad [e_2,e_3] = e_1, \qquad [e_3,e_1] = e_2.$$

In what follows, unless otherwise stated, we will always assume that the space $\mathbb{R}^n$ is oriented so that the basis $(e_1,\dots,e_n)$ has the positive orientation.

A vector space endowed with a bilinear anticommutative operation satisfying the Jacobi identity is called a *Lie algebra.* The space $\mathbb{R}^3$ with vector product gives us the first example of a Lie algebra.

For the Frenet frame of a curve in $\mathbb{R}^3$ to exist, it is not enough to assume that the curve is regular. We have to require the curve $r(l)$ to be *biregular*, i.e., such that its acceleration vector does not vanish anywhere.

Then the Frenet frame $(v,n,b)$ is

$$v = \frac{dr}{dl}, \quad n = \frac{d^2r}{dl^2}\Big/\left|\frac{d^2r}{dl^2}\right|, \quad b = [v,n].$$

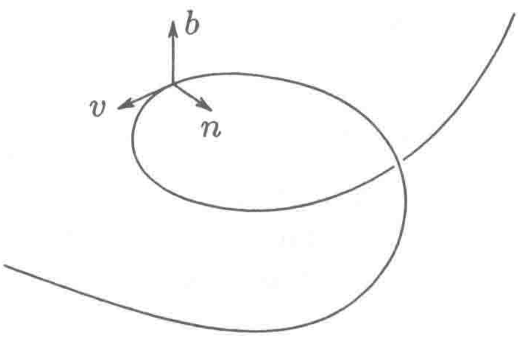


**Figure 1.10.** The Frenet frame for a curve in $\mathbb{R}^3$.

The vector $n$ is called the *principal normal* to the curve, and $b$ is called the *binormal.*

By definition,

$$\frac{dv}{dl} = kn \quad \text{and} \quad k > 0.$$

The Frenet frame is orthonormal and positively oriented. The quantity $k$ is called the *curvature* of a curve in $\mathbb{R}^3$. It may be equal to zero, so it is defined not only for biregular curves. It has a simple physical meaning: if we view our curve as the trajectory of a point-particle and take time for the natural parameter, then the curvature equals the length of the acceleration vector. The reciprocal quantity $R = \frac{1}{k}$ is the *radius of curvature* for the curve. The deformations of Frenet frames are described by the *Frenet equations* for a curve in $\mathbb{R}^3$.

**Theorem 1.11.** *The Frenet frame of a biregular curve in* $\mathbb{R}^3$ *satisfies the Frenet equations for a space curve:*

$$\frac{d}{dl}\begin{pmatrix} v \\ n \\ b \end{pmatrix} = \begin{pmatrix} 0 & k & 0 \\ -k & 0 & \kappa \\ 0 & -\kappa & 0 \end{pmatrix} \begin{pmatrix} v \\ n \\ b \end{pmatrix}. \tag{1.6}$$

**Proof.** As in the proof of Theorem 1.9, observe that $(v(l), n(l), b(l))$ for any $l$ form an orthonormal basis. In particular, the lengths of these vectors are kept constant, hence the derivative of each of these vectors is orthogonal to the vector itself:

$$\frac{d}{dl}\begin{pmatrix} v \\ n \\ b \end{pmatrix} = \begin{pmatrix} 0 & a_{12} & a_{13} \\ a_{21} & 0 & a_{23} \\ a_{31} & a_{32} & 0 \end{pmatrix} \begin{pmatrix} v \\ n \\ b \end{pmatrix}.$$

It remains to note that

$$\frac{d\langle v, n\rangle}{dl} = a_{12} + a_{21} = 0, \qquad \frac{d\langle n, b\rangle}{dl} = a_{23} + a_{32} = 0, \qquad \frac{d\langle v, b\rangle}{dl} = a_{13} + a_{31} = 0.$$

By the definition of curvature, we have $a_{12} = k$ and $a_{13} = 0$. Putting $\kappa = a_{23}$, we obtain equations (1.6). □

In fact, we restated the proof of Theorem 1.6 for $n = 3$. Namely, let $B(l)$ be a one-parameter family of orthogonal matrices specifying the transformations such that $B(l)e_1 = v(l)$, $B(l)e_2 = n(l)$, $B(l)e_3 = b(l)$, where $(e_1, e_2, e_3)$ is a fixed orthonormal positively oriented basis.

By Theorem 1.6 we have

$$\frac{dB}{dl}(l) = A(l)B(l),$$

where the matrix $A$ is skew-symmetric and by its construction

$$\frac{d}{dl}\begin{pmatrix} v \\ n \\ b \end{pmatrix} = A \begin{pmatrix} v \\ n \\ b \end{pmatrix}.$$

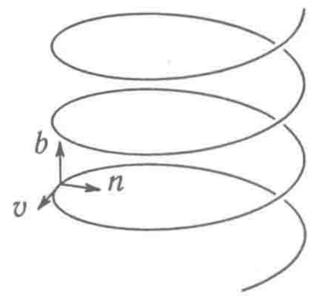


**Figure 1.11.** A spiral with the Frenet frame.

The quantity $\kappa$ involved in the Frenet equations is called the *torsion* of the curve.

We can state the following analog of Theorem 1.10.

**Theorem 1.12.** 1. *Let $k$ and $\kappa$ be smooth functions on an interval $[0, L]$ with $k > 0$ everywhere on this interval. Then there exists a smooth naturally parametrized curve $r\colon [0, L] \to \mathbb{R}^3$ with curvature $k(l)$ and torsion $\kappa(l)$. This curve may be found from the Frenet equations.*

2. *If two naturally parametrized curves $r_1(l)$ and $r_2(l)$ in $\mathbb{R}^3$ have the same curvatures and torsions, then there exists a proper motion $\varphi\colon \mathbb{R}^3 \to \mathbb{R}^3$ that maps one of the curves into the other: $\varphi(r_1(l)) = r_2(l)$.*

Since the proof of this theorem is quite similar to that of Theorem 1.10, we omit it.

If we consider a curve lying entirely in the plane $\mathbb{R}^2 \subset \mathbb{R}^3$, then the binormal to this curve will coincide with the binormal to the plane, so that

it will be constant. It follows from the Frenet equations that in this case $\kappa = 0$.

The converse is also true. If a curve $r(l)$ has torsion $\kappa = 0$, then by Theorem 1.12, for a given $k$, we can construct a planar curve $\tilde{r}(l)$ with curvature $k$. Since the curves $r(l)$ and $\tilde{r}(l)$ have the same curvatures and torsions, they can be mapped into one another by a proper motion of $\mathbb{R}^3$. Therefore $r(l)$ is a planar curve.

The equations $k = k(l)$ and $\kappa = \kappa(l)$, where $k$ and $\kappa$ are the curvature and torsion regarded as functions of the natural parameter $l$, are called the *natural equations of the curve*. By the previous theorem, they determine the curve uniquely up to motions of the entire space.

## Exercises to Chapter 1

1. Prove that the Jacobi matrix of the composition of two smooth mappings is the product of the Jacobi matrices of these mappings.

2. Prove that the rank of a Jacobi matrix does not depend on the choice of local coordinates.

3. Write down the Laplace operators on the plane,

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},$$

and in the 3-dimensional space,

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

in polar, spherical, and cylindrical coordinates.

4. Determine the functions $r = r(\varphi)$ that specify straight lines in polar coordinates.

5. Show that the length of a curve equals the limit of the lengths of polygonal lines consisting of segments that consecutively join finitely many points of the curve as the maximal length of the segments tends to zero.

6. Prove the following formula for the curvature of a planar curve $r(t)$ (with arbitrary parameter $t$):

$$|k| = \frac{|[\dot{r}, \ddot{r}]|}{|\dot{r}|^3}.$$

7. Suppose a half-line $OA$ revolves about the point $O$ with constant angular velocity $\omega$. Derive the equation for the trajectory of the point $P$ that

a) moves away from the point $O$ along the line with constant velocity (*spiral of Archimedes*);

b) moves along the line with velocity proportional to the distance $|OP|$ (*logarithmic spiral*).

Find the curvatures of the corresponding curves.

8. Prove that for any tangent to the curve $x^{2/3}+y^{2/3}=a^{2/3}$, the segment lying between the Euclidean coordinate axes has the constant length $a$.

9. Prove that for any tangent to the curve

$$y = \frac{a}{2}\log\frac{a+\sqrt{a^2-x^2}}{a-\sqrt{a^2-x^2}} - \sqrt{a^2-x^2},$$

the segment between the $y$-axis and the point of tangency has length $a$.

10. Let a circle of radius $a$ be rolling along a straight line without gliding. Derive the equation for the trajectory of a point fixed to the circle at a distance $d$ from its center. Find the curvatures of these curves.

11. Let a circle of radius $r$ be rolling without gliding over a circle of radius $R$ staying outside the circle. Derive the equation for the trajectory of a point of the rolling circle and find its curvature.

12. Find the curvature of the ellipse $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ at its apexes.

13. Find the curvature of the curve specified by the equation $F(x,y)=0$.

14. Find the curvature of the curve satisfying the differential equation $P(x,y)\,dx+Q(x,y)\,dy=0$.

15. Suppose a curve on a sphere intersects all the meridians at a given angle (*loxodrome*). Derive its equation and find the curvature and torsion.

16. Prove that if all the normal planes to a curve in $\mathbb{R}^3$ contain a fixed vector $\xi$, then the curve lies in a plane.

17. Prove that if all the normal planes to a curve pass through a fixed point $P$, then the curve lies on a sphere with center at this point.

18. Prove that a curve of identically zero curvature $k(l)\equiv 0$ is a straight line.

19. Prove that a curve with identically zero torsion $\kappa(l)\equiv 0$ lies in a plane and find this plane.

20. The *mixed product* of three vectors in $\mathbb{R}^3$ is defined as $(u,v,w) = ([u,v],w)$, i.e., as the scalar product of $[u,v]$ and $w$. Prove that a curve $r(t)$ is planar if and only if the mixed product of the vectors $\dot r, \ddot r$, and $\dddot r$ is equal to zero:

$$(\dot r, \ddot r, \dddot r) = 0.$$

21. Prove that the torsion of a curve $r(t)$ in $\mathbb{R}^3$ is equal to

$$\kappa = \frac{(\dot r, \ddot r, \dddot r)}{|[\dot r, \ddot r]|^2}.$$

22. Denote by $S$ the area between a planar curve and its secant passing parallel to a tangent at a distance $h$ from it. Express

$$\lim_{h\to 0} \frac{S^2}{h^3}$$

in terms of the curvature of the curve.

23. Prove that for any smooth closed curve $r = r(t)$ in $\mathbb{R}^3$ the following equality holds:

$$\int (r\, dk + \kappa b\, dl) = 0.$$

24. Solve the equation $r' = [\omega, r]$, $\omega = \text{const}$.

25. Prove that a curve lies on a sphere of radius $r$ if and only if

$$r^2 = \frac{1}{k^2}\left(1 + \frac{(k')^2}{(\kappa k)^2}\right)$$

(with the derivative taken with respect to the natural parameter).

26. Prove that if a curve lies on a sphere and has a constant curvature, then this is a circle.

27. For a smooth curve $r = r(l)$, let $n(l)$ be the curve formed by the tips of the normal vectors. Let $l^*$ be the natural parameter on this curve. Prove that

$$\frac{dl^*}{dl} = \sqrt{k^2 + \kappa^2}.$$

*Chapter 2*

# Symplectic and Pseudo-Euclidean Spaces

## 2.1. Geometric structures in linear spaces

**2.1.1. Pseudo-Euclidean and symplectic spaces.** In many applications we need a more general notion of scalar product than the Euclidean scalar product.

A *scalar product* on a finite-dimensional vector space $V$ is a function $\langle \xi, \eta \rangle$ of a pair of vectors $\xi, \eta \in V$ satisfying the following bilinearity condition:

$$\langle \lambda_1 \xi_1 + \lambda_2 \xi_2, \eta \rangle = \lambda_1 \langle \xi_1, \eta \rangle + \lambda_2 \langle \xi_2, \eta \rangle,$$
$$\langle \eta, \lambda_1 \xi_1 + \lambda_2 \xi_2 \rangle = \lambda_1 \langle \eta, \xi_1 \rangle + \lambda_2 \langle \eta, \xi_2 \rangle$$

for any $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\xi_1, \xi_2, \eta \in V$. This product is said to be *nondegenerate* if for each vector $\xi \neq 0$ there exists a vector $\eta$ such that $\langle \xi, \eta \rangle \neq 0$.

If $e_1, \dots, e_n$ is a basis in $V$, then bilinearity implies that

$$\langle \xi, \eta \rangle = \sum_{i,j}^{n} \xi^i \eta^j \langle e_i, e_j \rangle,$$

where $\xi = \xi^i e_i$, $\eta = \eta^j e_j$. Therefore, the product is specified by the Gram matrix $G = (g_{ij})$, where

$$g_{ij} = \langle e_i, e_j \rangle.$$

**Lemma 2.1.** *A scalar product is nondegenerate if and only if* $\det G \neq 0$.

**Proof.** A scalar product is degenerate if there exists a nonzero vector $\xi$ such that $\langle \xi, \eta \rangle = \xi^i \eta^j g_{ij} = 0$ for all $\eta$. In this case $g_{ij}\xi^i = 0$ for any $j$. This system of equalities may be rewritten as a matrix equation which has a nonzero solution if and only if $\det G = 0$. □

Of particular interest are the following two types of nondegenerate scalar products, namely, pseudo-Euclidean and symplectic.

a) A nondegenerate scalar product is said to be *pseudo-Euclidean* if it is *symmetric*,

$$\langle \xi, \eta \rangle = \langle \eta, \xi \rangle \quad \text{for all } \xi, \eta \in V,$$

but not positive definite.

b) A nondegenerate scalar product is called *symplectic* if it is *skew-symmetric*,

$$\langle \xi, \eta \rangle = -\langle \eta, \xi \rangle \quad \text{for all } \xi, \eta \in V.$$

Symmetric scalar products are also called *quadratic forms*. In linear algebra and number theory, by quadratic forms are also meant functions $q(x) = \langle x, x \rangle$, where $\langle \xi, \eta \rangle$ is a symmetric scalar product.

Symmetric and skew-symmetric scalar products can be reduced to a canonical form in special bases.

**Theorem 2.1.** 1. *If $\langle \cdot, \cdot \rangle$ is a symmetric scalar product on $V$, then there exists a basis $e_1, \ldots, e_n$ in $V$ such that*

$$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{for } i = j,\ 1 \le i \le p, \\ -1 & \text{for } i = j,\ p+1 \le i \le p+q \le n, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

2. *If $\langle \cdot, \cdot \rangle$ is a skew-symmetric scalar product on $V$, then there exists a basis $e_1, \ldots, e_n$ in $V$ such that*

$$\begin{cases} \langle e_{2i-1}, e_{2i} \rangle = -\langle e_{2i}, e_{2i-1} \rangle = 1 & \text{for } i = 1, \ldots, p,\ p \le [n/2], \\ \langle e_i, e_j \rangle = 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

REMARK. In the basis (2.2) the Gram matrix has the form

$$J = \begin{pmatrix} 0 & 1 & & & & & 0 \\ -1 & 0 & & & & & \\ & & \ddots & & & & \\ & & & 0 & 1 & & \\ & & & -1 & 0 & & \\ & & & & & 0 & \\ & & & & & & \ddots \\ 0 & & & & & & 0 \end{pmatrix}.$$

In the symplectic case with $\dim V = 2p$ we can introduce the following basis $\tilde{e}_1, \ldots, \tilde{e}_{2p}$:

$$e_1 = \tilde{e}_1, \quad e_2 = \tilde{e}_{p+1}, \ \ldots, e_{2j-1} = \tilde{e}_j, \quad e_{2j} = \tilde{e}_{p+j}, \quad j = 1, \ldots, p. \tag{2.3}$$

This basis is called *symplectic*, and in this basis the Gram matrix becomes

$$\tilde{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.4}$$

where 1 denotes the $p \times p$ identity matrix.

**Proof of Theorem 2.1.** 1. If $\langle \xi, \xi \rangle = 0$ for all vectors $\xi \in V$, then

$$\langle \xi + \eta, \xi + \eta \rangle = \langle \xi, \xi \rangle + \langle \eta, \eta \rangle + \langle \xi, \eta \rangle + \langle \eta, \xi \rangle = 2\langle \xi, \eta \rangle + \langle \xi, \xi \rangle + \langle \eta, \eta \rangle = 0.$$

Combined with condition $\langle \xi, \xi \rangle = \langle \eta, \eta \rangle = 0$ this implies the equality $\langle \xi, \eta \rangle = 0$. Therefore, the form $\langle \xi, \eta \rangle$ is identically zero, $p = q = 0$, and the basis $e_1, \ldots, e_n$ may be chosen arbitrarily.

If $\langle \xi, \xi \rangle \neq 0$ for some vector $\xi$, then we set $e_1 = \xi / \sqrt{|\langle \xi, \xi \rangle|}$ and denote by $V_1$ and $V_1^\perp$ the one-dimensional subspace spanned by $e_1$ and its orthogonal complement. Since $V_1^\perp$ is determined by the nontrivial equation

$$\langle \xi, x \rangle = 0 \quad \text{for } x \in V_1^\perp,$$

we see that $V_1^\perp$ is a hyperplane in $V$. Now we restrict $\langle \cdot, \cdot \rangle$ to $V_1^\perp$ and repeat the same argument. Then after finitely many steps we construct the required basis $e_1, e_2, \ldots, e_n$ (apart from ordering its vectors). This proves part 1 of the theorem.

2. If $\langle \xi, \eta \rangle = 0$ for any pair of vectors $\xi, \eta$, then the form $\langle \cdot, \cdot \rangle$ is identically zero, and the basis $e_1, \ldots, e_n$ can be chosen arbitrarily.

Otherwise, take $\xi$ and $\eta$ such that $\langle \xi, \eta \rangle \neq 0$. Set

$$e_1 = \xi, \quad e_2 = \frac{\eta}{\langle \xi, \eta \rangle}.$$

Denote by $V_1$ the subspace spanned by $e_1$ and $e_2$, and by $V_1^\perp$ its orthogonal complement. The pair of linearly independent equations

$$\langle \xi, x \rangle = \langle \eta, x \rangle = 0$$

determines $V_1^\perp$; hence the subspace $V_1^\perp$ has codimension 2 in $V$, i.e., $\dim V - \dim V_1^\perp = 2$.

Restrict $\langle \cdot, \cdot \rangle$ to $V_1^\perp$ and apply the same arguments. In this way, after finitely many steps, we construct a collection of vectors $e_1, \ldots, e_{2p}$ such that the form is identically zero on the orthogonal complement to them, say, $V_0$.

Now we supplement the collection $e_1, \ldots, e_{2p}$ with an arbitrary basis in $V_0$ to become a basis in $V$. This yields the required basis. $\square$

In the case of a symmetric scalar product (quadratic form), the tuple $(p, q, n-(p+q))$ is invariant, i.e., it depends on the form but not on the choice of the basis $e_1, \dots, e_n$ in which the Gram matrix takes the diagonal form. The sum $p+q$ is called the *rank* of the quadratic form, and the difference $p-q$, its *signature.*

If $p+q=n$ and $p>0$, $q>0$, the scalar product is called pseudo-Euclidean. A Cartesian space with such a scalar product is referred to as *pseudo-Euclidean* and is denoted by $\mathbb{R}^{p,q}$. The coordinates in which the scalar product has the form (2.1) are called pseudo-Euclidean.

A Cartesian space with a symplectic scalar product is called *symplectic*, and the coordinates in which the scalar product has the form (2.4) are called symplectic.

Theorem 2.1 implies the following statement.

**Corollary 2.1.** *A symplectic space has an even dimension.*

In what follows, when speaking about pseudo-Euclidean or symplectic spaces, we will assume that the coordinates chosen in these spaces are pseudo-Euclidean or symplectic, respectively.

**Lemma 2.2.** *A linear transformation $A\colon V \to V$ preserves a nondegenerate scalar product with Gram matrix $G$ if and only if*

$$A^\top G A = G.$$

**Proof.** Write down the equality $\langle \xi, \eta \rangle = \langle A\xi, A\eta \rangle$ in the coordinate form:

$$g_{ij}\xi^i\eta^j = g_{kl}a_i^k\xi^i a_j^l\eta^j.$$

This equality is fulfilled for all $\xi$ and $\eta$ if and only if

$$g_{ij} = a_i^k g_{kl} a_j^l,$$

i.e., for $G = A^\top G A$. □

Since $\det G \neq 0$ for any nondegenerate scalar product, the equality $\det A^\top = \det A$ implies the following result.

**Corollary 2.2.** *If a linear transformation $A$ preserves a nondegenerate scalar product, then it is invertible and* $\det A = \pm 1$.

Indeed, $\det G = \det A^\top \det G \det A = \det^2 A \det G$, which implies the equality $\det^2 A = 1$.

**Theorem 2.2.** *Linear transformations $A\colon V \to V$ preserving a nondegenerate scalar product form a group.*

**Proof.** Obviously, the composition of such transformations preserves the scalar product. Let a transformation $A$ preserve the scalar product. As we have just shown, it is invertible. Multiplying both sides of the equality $A^\top GA = G$ by $A^{-1}$ on the right and by $(A^{-1})^\top$ on the left we obtain

$$G = (A^{-1})^\top GA^{-1}.$$

Hence the linear transformation $A^{-1}$ also preserves the scalar product. □

The group of linear transformations of the space $\mathbb{R}^{p,q}$ preserving the pseudo-Euclidean scalar product is denoted by $\mathrm{O}(p,q)$. Specifying pseudo-Euclidean coordinates we obtain that this group is a group of matrices preserving the scalar product

$$\langle \xi, \eta \rangle = \xi^1\eta^1 + \cdots + \xi^p\eta^p - \xi^{p+1}\eta^{p+1} - \cdots - \xi^{p+q}\eta^{p+q}.$$

The subgroup of $\mathrm{O}(p,q)$ formed by the matrices with determinant equal to 1 is denoted by $\mathrm{SO}(p,q)$.

In the next section we will consider these groups and pseudo-Euclidean spaces in more detail.

**2.1.2. Symplectic transformations.** The group of linear transformations of the space $\mathbb{R}^{2n}$ preserving a symplectic scalar product is denoted by $\mathrm{Sp}(n,\mathbb{R})$ and is called a *symplectic group*. This is the group of matrices preserving the scalar product in the basis (2.3):

$$\langle \xi, \eta \rangle = \xi^1\eta^{n+1} + \cdots + \xi^n\eta^{2n} - \xi^{n+1}\eta^1 - \cdots - \xi^{2n}\eta^n.$$

EXAMPLES. 1. Let $n = 1$. A matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ belongs to $\mathrm{Sp}(1,\mathbb{R})$ if and only if

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The left-hand side of this equation equals

$$\begin{pmatrix} 0 & ad - bc \\ -(ad - bc) & 0 \end{pmatrix}.$$

Therefore, $\mathrm{Sp}(1,\mathbb{R})$ consists exactly of $(2 \times 2)$ matrices with determinant equal to 1,

$$\mathrm{Sp}(1,\mathbb{R}) = \mathrm{SL}(2).$$

2. Consider a symplectic basis $\tilde{e}_1, \ldots, \tilde{e}_{2n}$ (i.e., a basis with Gram matrix (2.4)). Then any linear transformation $B$ of the $n$-dimensional space with basis $\tilde{e}_1, \ldots, \tilde{e}_n$ induces the following symplectic transformation of the $2n$-dimensional space:

$$A = \begin{pmatrix} B & 0 \\ 0 & (B^\top)^{-1} \end{pmatrix}, \quad B \in \mathrm{GL}(n). \tag{2.5}$$

**Lemma 2.3.** *Let $V$ be a $2n$-dimensional symplectic space, let $A \in \mathrm{Sp}(n, \mathbb{R})$, and let $\mu$ be a real eigenvalue of $A$. Then the characteristic polynomial $p(\lambda) = \det(A - \lambda \cdot 1)$ of the transformation $A$ has the form*

$$p(\lambda) = (\lambda - \mu)\Big(\lambda - \frac{1}{\mu}\Big)p'(\lambda),$$

*where $p'(\lambda)$ is the characteristic polynomial of a matrix $A' \in \mathrm{Sp}(n-1, \mathbb{R})$.*

**Proof.** It is seen from the proof of Theorem 2.1 that in the construction of a basis with property (2.2) the vector $e_1$ can be chosen arbitrarily. Let us construct a basis as in the proof of Theorem 2.1, with $e_1$ such that $Ae_1 = \mu e_1$. It is easily seen that

$$Ae_2 = \frac{1}{\mu}\, e_2 + \sum_{i \neq 2} b^i e_i,$$

since $\langle Ae_1, Ae_2 \rangle = 1 = \langle e_1, e_2 \rangle$. Hence the transformation $A$ is given by the matrix

$$A = \begin{pmatrix} \mu & * & * \\ 0 & \frac{1}{\mu} & 0 \\ 0 & * & A' \end{pmatrix},$$

where $A'$ specifies the symplectic transformation. Therefore, we have $p(\lambda) = \det(A - \lambda \cdot 1) = (\lambda - \mu)\big(\lambda - \frac{1}{\mu}\big)\det(A' - \lambda \cdot 1)$. □

Now we can prove the following statement.

**Lemma 2.4.** *If $A \in \mathrm{Sp}(n, \mathbb{R})$, then* $\det A = 1$.

**Proof.** The determinant of a matrix $A$ is equal to the product of the roots of its characteristic polynomial $p(\lambda) = \det(A - \lambda \cdot 1)$. Lemma 2.3 implies that

$$\det(A - \lambda \cdot 1) = (\lambda - \mu_1)\Big(\lambda - \frac{1}{\mu_1}\Big)\cdots(\lambda - \mu_k)\Big(\lambda - \frac{1}{\mu_k}\Big)\tilde{p}(\lambda),$$

where the polynomial $\tilde{p}(\lambda)$ is either identically equal to one or has no real roots. In the former case $\det A = \mu_1\mu_1^{-1} \cdots \mu_k\mu_k^{-1} = 1$. In the latter case the roots of the real polynomial $\tilde{p}(\lambda)$ are of the form $\nu_1, \bar{\nu}_1, \ldots, \nu_l, \bar{\nu}_l$, so that

$$\det(A) = p(0) = \Big(\prod_{j=1}^{k} \mu_j\mu_j^{-1}\Big) \times \Big(\prod_{j=1}^{l} \nu_j\bar{\nu}_j\Big) > 0.$$

Since $\det A = \pm 1$, this implies that $\det A = 1$. □

Now that we know that $\det A = 1$, we will show that the characteristic polynomial of a symplectic matrix has a stronger property.

**Lemma 2.5.** *If $A \in \mathrm{Sp}(n,\mathbb{R})$, then the characteristic polynomial $p(\lambda) = \det(A - \lambda \cdot 1)$ is reciprocal to itself*:

$$p(\lambda) = \lambda^{2n} p\Big(\frac{1}{\lambda}\Big).$$

**Proof.** Denote by $J \in \mathrm{Sp}(n,\mathbb{R})$ the Gram matrix,

$$J = \begin{pmatrix} 0 & 1 & & & 0 \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ 0 & & & -1 & 0 \end{pmatrix}.$$

Observe that $J^2 = -1$ and $\det J = 1$. The symplecticity condition is

$$A^\top J A = J.$$

This implies that

$$\begin{aligned} p(\lambda) = \det(A - \lambda \cdot 1) &= \det J \, \det(A - \lambda \cdot 1) = \det(JA - \lambda J) \\ &= \det(JA - \lambda A^\top JA) = \det(1 - \lambda A^\top) \det JA \\ &= \det(1 - \lambda A^\top) = \lambda^{2n} \det(\lambda^{-1} \cdot 1 - A)^\top \\ &= \lambda^{2n} \det(\lambda^{-1} \cdot 1 - A). \end{aligned}$$

But since $A$ is a $2n \times 2n$ matrix, we conclude that $\det(\lambda^{-1} \cdot 1 - A) = (-1)^{2n} \det(A - \lambda^{-1} \cdot 1)$; hence

$$p(\lambda) = \lambda^{2n} \det(A - \lambda^{-1} \cdot 1) = \lambda^{2n} p\Big(\frac{1}{\lambda}\Big).$$

□

These lemmas entail an important property of the *spectrum* of a symplectic matrix, i.e., the set of the roots of its characteristic polynomial.

**Corollary 2.3.** *The spectrum of a matrix $A \in \mathrm{Sp}(n,\mathbb{R})$ is symmetric about the real axis and the circle $|\lambda| = 1$. It falls into*:

1) *4-tuples $\lambda$, $\bar\lambda$, $\lambda^{-1}$, $\bar\lambda^{-1}$, where $|\lambda| \neq 1$ and $\mathrm{Im}\,\lambda \neq 0$*;
2) *pairs $\lambda$, $\lambda^{-1}$ lying on the real axis ($\bar\lambda = \lambda$)*;
3) *pairs $\lambda$, $\lambda^{-1}$ lying on the circle $|\lambda| = 1$ (so that $\lambda^{-1} = \bar\lambda$).*

A specific feature of a symplectic space is dissimilarity between hyperplanes. For example, in $\mathbb{R}^4$ no symplectic transformation maps the plane $x^2 = x^4 = 0$ into the plane $x^3 = x^4 = 0$. Indeed, any two vectors $\xi$ and $\eta$ of the plane $x^3 = x^4 = 0$ satisfy the equality $\langle \xi, \eta \rangle = 0$, whereas for the vectors $\xi = (1,0,0,0)$ and $\eta = (0,0,1,0)$ of the plane $x^2 = x^4 = 0$ we have $\langle \xi, \eta \rangle = 1$.

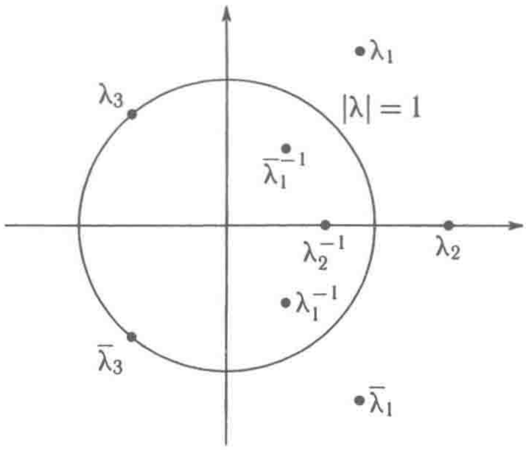


**Figure 2.1.** Roots of the characteristic polynomial of a symplectic matrix.

A subspace $V$ of a $2n$-dimensional symplectic space is said to be *Lagrangian* if it is $n$-dimensional and

$$\langle \xi, \eta \rangle = 0$$

for any two vectors $\xi, \eta \in V$.

All coordinate Lagrangian subspaces have the form

$$x^{i_1} = \cdots = x^{i_n} = 0, \tag{2.6}$$

where $\langle e_{i_j}, e_{i_k} \rangle = 0$.

Two subspaces $L_1$ and $L_2$ in $\mathbb{R}^N$ are *transversal* to each other if their vectors span the entire space $\mathbb{R}^N$. If $\dim L_1 + \dim L_2 = N = \dim \mathbb{R}^N$, then transversality implies that $L_1$ and $L_2$ intersect only at the origin.

To each coordinate Lagrangian subspace $L$ of the form (2.6) corresponds the transversal coordinate Lagrangian subspace $L'$. It is specified by the equations

$$x^{j_1} = \cdots = x^{j_n} = 0,$$

where $j_l \neq i_k$ for all $l, k = 1, \ldots, n$.

**Lemma 2.6.** *Let $\widetilde{L}$ be a Lagrangian subspace in $\mathbb{R}^{2n}$. Then there exists a coordinate Lagrangian subspace $L$ transversal to $\widetilde{L}$.*

**Proof.** Making use of transformation (2.5), we construct a new symplectic basis such that its first $k$ vectors $\tilde{e}_1, \ldots, \tilde{e}_k$ span the entire intersection of $\widetilde{L}$ with the subspace $(\tilde{x}^{n+1} = 0, \ldots, \tilde{x}^{2n} = 0)$.

Define the subspace $L$ transversal to $\widetilde{L}$ as the one spanned by the vectors

$$\tilde{e}_{k+1}, \ldots, \tilde{e}_n, \tilde{e}_{n+1}, \ldots, \tilde{e}_{n+k}.$$

Obviously, this subspace is Lagrangian. Consider the vector $\eta \in L \cap \widetilde{L}$. Since $\langle \eta, L\rangle = \langle \eta, \widetilde{L}\rangle = 0$, we have

$$\langle \eta, \tilde{e}_j\rangle = 0, \quad j = 1, \dots, n.$$

Therefore, $\eta = \sum_{j=1}^{n} \eta^j \tilde{e}_j$ and hence $\eta = 0$. □

This lemma implies that each Lagrangian plane is representable as the graph of a mapping of a coordinate Lagrangian subspace. Namely, the following theorem holds.

**Theorem 2.3.** *For any pair of transversal Lagrangian subspaces* $(L, \widetilde{L})$, *one can construct a symplectic basis of the form* (2.3) *such that*

$$\tilde{e}_1, \dots, \tilde{e}_n \in \widetilde{L}, \quad \tilde{e}_{n+1}, \dots, \tilde{e}_{2n} \in L,$$

*and the Gram matrix has the form*

$$\langle \tilde{e}_i, \tilde{e}_{n+j}\rangle = \delta_{ij}, \quad i, j = 1, \dots, n$$

(*with all other elements equal to zero*).

The proof is obvious.

## 2.2. The Minkowski space

**2.2.1. The event space of the special relativity theory.** The most important class of pseudo-Euclidean spaces is the class of spaces $\mathbb{R}^{1,n}$. In this case the scalar product in pseudo-Euclidean coordinates $x^0, x^1, \dots, x^n$ becomes

$$\langle \xi, \eta\rangle = \xi^0\eta^0 - \xi^1\eta^1 - \dots - \xi^n\eta^n.$$

It specifies the *Minkowski metric.*

For $n = 3$ we obtain the *Minkowski space,* which is the 4-dimensional space-time of the special relativity theory. In this case $x^0 = ct$, where $c$ is the velocity of light in vacuum and $t$ is time, while $x^1, x^2, x^3$ are coordinates in the three-dimensional physical space. The points of the space $\mathbb{R}^{1,3}$ are instantaneous events, and the quantity

$$\sqrt{\langle x - y, x - y\rangle}$$

for $x, y \in \mathbb{R}^{1,3}$ is the *space-time interval* between the points (events) $x$ and $y$.

The squared length of a vector may be positive, negative, or zero. A vector $\xi$ is a *light-like vector* if $\langle \xi, \xi\rangle = 0$. If $\langle \xi, \xi\rangle > 0$, the vector $\xi$ is said to be *time-like,* and when $\langle \xi, \xi\rangle < 0$, it is called *space-like*.

This terminology has physical grounds. Namely, the special relativity theory relies on the following postulates:

1) A material point-particle cannot move in the space at a velocity greater than the velocity of light. For a particle with mass, the velocity

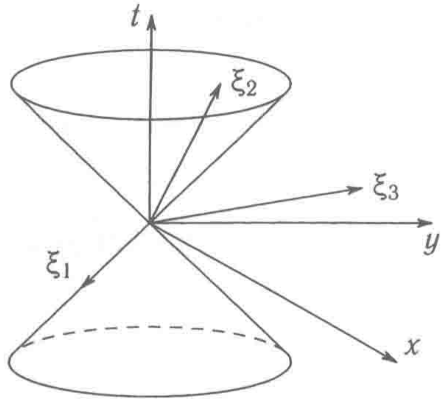


**Figure 2.2.** Light-, time- and space-like vectors; light cone.

is always less than the velocity of light, while the velocity of a zero-mass particle equals the velocity of light.

2) In all the inertial frames the velocity of light has a constant value

$$c \approx 2.998 \cdot 10^{10} \text{ cm/sec}.$$

The curves $r(\tau)$ in $\mathbb{R}^{1,3}$ with velocity vectors of zero length, i.e.,

$$\langle r_\tau, r_\tau \rangle = 0, \quad \text{where } r_\tau = \frac{dr}{d\tau},$$

are the world lines of zero-mass particles, e.g., photons. Therefore, the light rays propagate along these lines.

If the velocity vector $r_\tau = (x^0_\tau, x^1_\tau, x^2_\tau, x^3_\tau)$ of a curve $r(\tau)$ is everywhere time-like, then $r(\tau)$ is a world line of a particle with mass. Consider such a world line $r(\tau)$. Its length in the Minkowski metric is

$$l = \int_a^b |r_\tau|\, d\tau = \int_a^b \sqrt{(x^0_\tau)^2 - \sum_{\alpha=1}^{3} (x^\alpha_\tau)^2}\, d\tau > 0.$$

The quantity $l/c$ is called the proper time elapsed for the particle. For a particle at rest ($x^1_\tau = x^2_\tau = x^3_\tau = 0$), it is equal to the interval of world time between the events $r(a)$ and $r(b)$, i.e., $x^0(b) - x^0(a)$.

EXAMPLE. For a particle at rest described by the world line

$$a \le x^0 \le b, \quad x^\alpha = 0, \quad \alpha = 1, 2, 3,$$

the time elapsed is $x^0(b) - x^0(a)$. Let the life of another particle be described by a world line $(x^0(\tau),\ x^\alpha(\tau))$, where $\tau = t = x^0/c$, $a \le x^0 \le b$, with the same initial and end points. Assume that this second particle moves during

this period of time: $x^\alpha(\tau) \neq 0$, $\alpha = 1, 2, 3$. Then its proper time is

$$\frac{l'}{c} = \int_a^b \frac{1}{c}\sqrt{c^2 - \sum_{\alpha=1}^{3}\Big(\frac{dx^\alpha}{dt}\Big)^2}\, dt.$$

Obviously, we have

$$l'/c < b - a = l/c.$$

Thus we see that *time contracts at movement.*

The physical principle of constant velocity of light is expressed by the requirement that the change from one inertial frame to another is determined by a motion of the Minkowski space, i.e., by a transformation $\varphi\colon \mathbb{R}^{1,3} \to \mathbb{R}^{1,3}$ preserving space-time intervals. This means that

$$\langle \varphi(x) - \varphi(y), \varphi(x) - \varphi(y)\rangle = \langle x - y, x - y\rangle$$

for each pair of points $x$ and $y$ in $\mathbb{R}^{1,3}$.

**Theorem 2.4.** *Any motion of the space $\mathbb{R}^{1,n}$ is an affine transformation.*

**Proof.** The proof differs very little from that of Theorem 1.4 for the Euclidean space. Let $\varphi$ be a motion of the space $\mathbb{R}^{1,n}$ and $\varphi(0) = b$. Since the shift $x \to x - b$ is a motion, define $\psi$ as the composition of $\varphi$ and this shift, $\psi\colon x \to \varphi(x) - b$. This is a motion with $\psi(0) = 0$.

Preservation of space-time intervals means that

$$\langle \psi(x) - \psi(y), \psi(x) - \psi(y)\rangle = \langle x - y, x - y\rangle \tag{2.7}$$

for $x, y \in \mathbb{R}^{1,n}$. In particular, for $y = 0$ we have

$$\langle \psi(x), \psi(x)\rangle = \langle x, x\rangle. \tag{2.8}$$

Rewriting (2.7) as

$$\langle \psi(x), \psi(x)\rangle - 2\langle \psi(x), \psi(y)\rangle + \langle \psi(y), \psi(y)\rangle = \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle,$$

and taking (2.8) into account, we obtain that

$$\langle \psi(x), \psi(y)\rangle = \langle x, y\rangle$$

for all $x, y \in \mathbb{R}^{1,n}$.

Let $e_0, \dots, e_n$ be the orthonormal basis in $\mathbb{R}^{1,n}$ that corresponds to the pseudo-Euclidean coordinates, and let $x_0, \dots, x_n$ be the points for which the basis vectors are radius-vectors. Denote by $\tilde{e}_i$ the radius-vector of the point $\psi(x_i)$. The basis $\tilde{e}_0, \dots, \tilde{e}_n$ is orthonormal. Consider the linear transformation $\chi\colon x \to Ax$ which maps $\tilde{e}_i$ into $e_i$, $i = 0, \dots, n$. Obviously, this transformation is a motion.

Therefore, the composition $\chi\psi$ is a motion that leaves the origin fixed and satisfies the equalities $\chi\psi(e_i) = e_i$ for $i = 0, 1, \ldots, n$. Any vector $x$ is expanded in this basis as

$$x = \langle x, e_0 \rangle e_0 - \sum_{i=1}^{n} \langle x, e_i \rangle e_i.$$

For any point $x \in \mathbb{R}^{1,n}$ we have

$$\langle x, e_i \rangle = \langle \chi\psi(x), e_i \rangle, \quad i = 0, \ldots, n.$$

Therefore, $\chi\psi(x) = x$, i.e., this is the identity transformation; hence $\psi = A^{-1}$, which is a linear transformation, and

$$\varphi(x) = A^{-1}x + b.$$

The proof is completed. □

The definition of a motion and the proof of the theorem can be extended to arbitrary spaces $\mathbb{R}^{p,q}$ with obvious changes. A similar result for Euclidean spaces was obtained in Theorem 1.4.

**Corollary 2.4.** *Every motion of the space $\mathbb{R}^{1,n}$ is uniquely representable in pseudo-Euclidean coordinates as*

$$x \to Ax + b,$$

*where*

$$A^{\top} \begin{pmatrix} 1 & & & 0 \\ & -1 & & \\ & & \ddots & \\ 0 & & & -1 \end{pmatrix} A = \begin{pmatrix} 1 & & & 0 \\ & -1 & & \\ & & \ddots & \\ 0 & & & -1 \end{pmatrix}. \tag{2.9}$$

**2.2.2. The Poincaré group.** The group of motions of $\mathbb{R}^{1,3}$ is called the *Poincaré group.*

Consider the subgroup of all motions of the subspace $\mathbb{R}^{1,1}$ in more detail. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The equation (2.9) becomes

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which can be rewritten as the system of equations

$$a^2 - c^2 = d^2 - b^2 = 1, \quad ab - cd = 0.$$

The solutions of this system fall into four families:

$$
\begin{aligned}
&\begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix}, \quad \begin{pmatrix} -\cosh\psi & -\sinh\psi \\ -\sinh\psi & -\cosh\psi \end{pmatrix}, \\
&\begin{pmatrix} \cosh\psi & -\sinh\psi \\ \sinh\psi & -\cosh\psi \end{pmatrix}, \quad \begin{pmatrix} -\cosh\psi & \sinh\psi \\ -\sinh\psi & \cosh\psi \end{pmatrix},
\end{aligned}
\tag{2.10}
$$

where $\psi$ is an arbitrary real parameter. The top two families consist of unimodular transformations ($\det A = 1$), and the bottom two, of transformations with $\det A = -1$. This implies the following result.

**Theorem 2.5.** *The group* $\mathrm{O}(1,n)$ *consists of four components.*

If $A \in \mathrm{O}(1,n)$, then $\langle Ae_0, e_0\rangle \neq 0$. Indeed, otherwise the vector $Ae_0$ would be a linear combination of $e_1, \dots, e_n$, implying that $\langle Ae_0, Ae_0\rangle < 0$. But this contradicts the fact that $\langle Ae_0, Ae_0\rangle = \langle e_0, e_0\rangle = 1$.

Let $\mathbb{Z}_2$ be the group formed by the numbers $\pm 1$ with the multiplication operation. Define the mapping $\varphi\colon \mathrm{O}(1,n) \to \mathbb{Z}_2 \times \mathbb{Z}_2$ by

$$\varphi(A) = (\det A, \operatorname{sgn}\langle Ae_0, e_0\rangle),$$

where

$$\operatorname{sgn}\lambda = \begin{cases} 1 & \text{for } \lambda > 0, \\ 0 & \text{for } \lambda = 0, \\ -1 & \text{for } \lambda < 0. \end{cases}$$

Consider the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ with componentwise multiplication

$$(a,b)\cdot(a',b') = (aa', bb').$$

**Lemma 2.7.** *The mapping* $\varphi$ *is a homomorphism of the group* $\mathrm{O}(1,n)$ *on the group* $\mathbb{Z}_2 \times \mathbb{Z}_2$.

**Proof.** The transformations in $\mathrm{O}(1,n)$ preserve the *light cone*

$$\langle \xi, \xi\rangle = 0,$$

which is the boundary of the domain formed by the vectors $\xi$ with $\langle\xi,\xi\rangle > 0$. This domain consists of two components, which are the interior parts of the cone. If $\operatorname{sgn}\langle Ae_0, e_0\rangle = 1$, each component is mapped into itself, and if $\operatorname{sgn}\langle Ae_0, e_0\rangle = -1$, they interchange. This implies that $\varphi$ is a homomorphism.

Now we list the values of $\varphi$ for the families in (2.10): $(1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$. Hence for $n = 1$ the image of $\varphi$ coincides with $\mathbb{Z}_2 \times \mathbb{Z}_2$. For $n > 1$ any transformation in $\mathrm{O}(1,1)$ may be extended in a trivial way ($Ae_2 = e_2, \dots, Ae_n = e_n$) to a transformation in $\mathrm{O}(1,n)$. Therefore, $\varphi$ is a homomorphism on the entire group $\mathbb{Z}_2 \times \mathbb{Z}_2$ for any $n$. $\square$

If $\operatorname{sgn}\langle Ae_0, e_0\rangle = 1$, then the transformation $A$ is called *orthochronous*, meaning that it does not change the direction of time. In pseudo-Euclidean geometry a transformation $A$ is said to be *proper* if $\varphi(A) = (1,1)$, i.e., it preserves the orientation of the space $\mathbb{R}^{1,3}$ and the direction of time.

**2.2.3. Lorentz transformations.** The transformations forming the group $O(1,3)$ appeared initially in the special relativity theory as *Lorentz transformations*. They describe the change from one inertial reference frame to another. Consider them in more detail.

In this subsection we will denote the space coordinates by $x, y, z$ as is customary in physics.

Denote by $ct, x, y, z$ the coordinates in a frame $K$ and by $ct'$, $x'$, $y'$, $z'$, the coordinates in a frame $K'$.

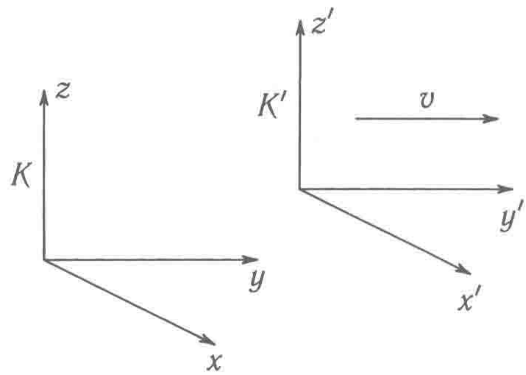


**Figure 2.3.** Different reference frames moving relative to each other.

Suppose that the origin of the frame $K'$, moving along the $x$-axis with constant velocity $v$, coincides with the origin of the frame $K$ at time $t = 0$. The transition formulas have the form

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix}$$

and are specified by a matrix of the group $O(1,3)$ because this transformation must preserve space-time intervals in $\mathbb{R}^{1,3}$. Naturally, this transformation must continuously depend on $v$ and become the identity transformation for $v = 0$. Hence it must belong to the identity component of the group $O(1,3)$ and have the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix}.$$

The point with $x' = y' = z' = 0$ corresponds to the origin of the frame $K'$, for which $x/t = v$, and we obtain

$$\tanh\psi = \frac{\sinh\psi}{\cosh\psi} = \frac{x}{ct} = \frac{v}{c},$$

$$\cosh\psi = \frac{1}{\sqrt{1-v^2/c^2}}, \quad \sinh\psi = \frac{v/c}{\sqrt{1-v^2/c^2}}.$$

The formulas of the *Lorentz transformation* become

$$\begin{aligned} t &= \frac{t' + x'v/c^2}{\sqrt{1-v^2/c^2}}, \\ x &= \frac{vt' + x'}{\sqrt{1-v^2/c^2}}, \quad y = y', \quad z = z'. \end{aligned} \tag{2.11}$$

As $v/c \to 0$, the Lorentz transformations turn into the *Galilean transformations*

$$t = t', \quad x = vt' + x',$$

which approximate the former with high accuracy if the velocity $v$ is much less than the velocity of light $c$, $v \ll c$.

Here we point out some physical phenomena which follow from the Lorentz transformations.

1. Contraction of length. Suppose we have a rod that is fixed relative to the frame $K$, is parallel to the $x$-axis, and has a length $l$ in this frame. The $x$-coordinates of its endpoints are $x_0$ and $x_0 + l$. The difference between the coordinates of its endpoints in the frame $K'$ is

$$l' = l\sqrt{1-v^2/c^2}.$$

Therefore, *the length of the rod contracts in a moving reference frame.*

2. Addition of velocities. Let $r(\tau)$ be the world line of a particle with mass. Then the tangent vector at each point of this line is time-like. We normalize it to be of unit length. Such a vector is called the *velocity 4-vector.* It is equal to

$$u = \left(\frac{1}{\sqrt{1-(v/c)^2}}, \frac{v}{c\sqrt{1-(v/c)^2}}\right), \tag{2.12}$$

where $v = \left(\frac{dx^1}{dt}, \frac{dx^2}{dt}, \frac{dx^3}{dt}\right)$ is the usual three-dimensional velocity vector.

When changing from one inertial frame to another the velocity 4-vector, like any vector in $\mathbb{R}^{1,3}$, transforms by the Lorentz formulas. Let us consider the consequences of this.

Let the frame $K'$ move relative to the frame $K$ at a constant velocity $w$ along the $x$-axis. Denote the three-dimensional velocity vectors of the

particle with mass by $v = (v_x, v_y, v_z)$ and $v' = (v'_x, v'_y, v'_z)$ as measured in the frames $K$ and $K'$ respectively, and similarly the velocity 4-vectors by $u$ and $u'$.

The Lorentz transformation (2.11) for 4-vectors $u = (u^0, u^1, u^2, u^3)$ becomes

$$u^0 = \frac{1}{\sqrt{1-(w/c)^2}}\Big(u'^0 + \frac{w}{c}u'^1\Big),$$
$$u^1 = \frac{1}{\sqrt{1-(w/c)^2}}\Big(\frac{w}{c}\,u'^0 + u'^1\Big), \quad u^2 = u'^2, \quad u^3 = u'^3.$$

Substituting here the expressions (2.12) for the components of the 4-vectors we derive the following formulas known as the *relativistic law of addition of velocities*:

$$v_x = \frac{v'_x + w}{1 + \frac{v'_x w}{c^2}}, \quad v_y = \frac{v'_y\sqrt{1-(w/c)^2}}{1 + \frac{v'_x w}{c^2}}, \quad v_z = \frac{v'_z\sqrt{1-(w/c)^2}}{1 + \frac{v'_x w}{c^2}}.$$

If the particle itself moves along the $x$-axis, these formulas take a particularly simple form:

$$v_x = \frac{v'_x + w}{1 + \frac{v'_x w}{c^2}},$$

with all other components of the three-dimensional vectors $v$ and $v'$ equal to zero.

If we substitute here $v'_x = c$, then for any $w$ we obtain $v_x = c$, which tells us in mathematical terms that the velocity of light is the same in all inertial frames. This fact was first observed in the Michelson–Morley experiment.

One more well-known physical phenomenon is contraction of the proper time of a moving particle, which was already stated in 2.2.1.

## Exercises to Chapter 2

1. Prove that the type (i.e., the pair $p, q$) of the pseudo-Euclidean scalar product does not depend on the choice of the basis.

2. Let $g_{ij} = -g_{ji}$ be a nonsingular skew-symmetric matrix specifying a symplectic product in $\mathbb{R}^{2n}$. Prove that a linear subspace of $\mathbb{R}^{2n}$ such that the symplectic product restricted to this subspace is identically zero is of dimension at most $n$.

3. Define the "vector product" of vectors $\xi = (\xi^0, \xi^1, \xi^2)$ and $\eta = (\eta^0, \eta^1, \eta^2)$ in the space $\mathbb{R}^{1,2}$ by the formula

$$\xi \times \eta = (\xi^1\eta^2 - \xi^2\eta^1, \xi^0\eta^2 - \xi^2\eta^0, \xi^1\eta^0 - \xi^0\eta^1),$$

where $\langle e_0, e_0 \rangle = 1$ and $\langle e_1, e_1 \rangle = \langle e_2, e_2 \rangle = -1$. Prove that

$$e_0 \times e_1 = -e_2, \quad e_0 \times e_2 = e_1, \quad e_1 \times e_2 = e_0,$$

the Jacobi identity holds,

$$\xi_1 \times (\xi_2 \times \xi_3) + \xi_3 \times (\xi_1 \times \xi_2) + \xi_2 \times (\xi_3 \times \xi_1) = 0,$$

and this vector product is invariant with respect to $\mathrm{SO}(1,2)$.

4. Let $r = r(l)$ be a time-like curve in $\mathbb{R}^{1,2}$ with natural parameter $l$ (i.e., $\dot{r}^2 = (\dot{r}^0)^2 - (\dot{r}^1)^2 - (\dot{r}^2)^2 \equiv 1$) such that $\dot{r}^0 > 0$. Set $v = \dot{r}$, $\dot{v} = kn$, $b = n \times v$. Prove the "Frenet formulas":

$$\dot{v} = kn, \quad \dot{n} = kv - \kappa b, \quad \dot{b} = \kappa n.$$

5. Solve the following equation in $\mathbb{R}^{1,2}$:

$$\dot{r} = \omega \times r, \quad \omega = \mathrm{const}\,.$$

*Chapter 3*

# Geometry of Two-Dimensional Manifolds

## 3.1. Surfaces in three-dimensional space

**3.1.1. Regular surfaces.** The existence of a natural parameter on a curve implies that each part of the curve can be mapped into the real line preserving the distances between the points. In contrast, two-dimensional surfaces in 3-dimensional space possess intrinsic geometry. In general, no neighborhood of a point on a surface may be mapped onto a domain of the Euclidean plane so as to preserve the distances between the points.

Before considering the geometry of surfaces, we will discuss how they can be specified. Following the tradition, we will denote coordinates in $\mathbb{R}^3$ by $x$, $y$, $z$.

The most usual way is to specify a surface as the graph of a function

$$z = f(x, y).$$

However, if we consider the unit sphere,

$$x^2 + y^2 + z^2 = 1,$$

it can be described by the graph of a function $f(x, y)$ in a neighborhood of any of its points except for those with $z = 0$. If we take a point of the sphere with $z = 0$, then one of the remaining coordinates, say, $x$, does not vanish at this point, so that in a neighborhood where $x \neq 0$ the sphere can

be represented by the graph

$$x = \tilde{f}(y, z) = \pm\sqrt{1 - y^2 - z^2}.$$

The sign in the right-hand side is taken to coincide with that of $x$. Similarly, in general a surface in $\mathbb{R}^3$ is defined as follows.

A set of points, $S \subset \mathbb{R}^3$, is a *regular surface* if, in a neighborhood of each of its points, it can be represented as the graph of a smooth function $z = f(x, y)$ in appropriate Cartesian coordinates $x$, $y$, $z$. The functions $f$ may be different for different points of $S$.

The graph representation of a surface is a particular case of the two other ways of specifying it:

a) as the set of zeros of a smooth function;

b) parametrically, i.e., as the image of a mapping $r\colon U \to \mathbb{R}^3$ of a domain $U \subset \mathbb{R}^2$.

Indeed, if a surface $S$ is the graph of a function $z = f(x, y)$ in a neighborhood of a point $(x_0, y_0, z_0) \in S$, then it is representable in the same neighborhood as the set of zeros of the function

$$F(x, y, z) = z - f(x, y)$$

and as the image of the following mapping of a neighborhood of the point $(x_0, y_0) \in \mathbb{R}^2$:

$$(x, y) \to r(x, y) = (x, y, f(x, y)).$$

All these ways of specifying a surface are equivalent. Before proving this, we recall the implicit function theorem.

**Theorem 3.1.** *Let $F\colon U \to \mathbb{R}^n$ be a smooth mapping of a domain $U$ of the space $\mathbb{R}^{n+k}$ with coordinates $(x^1, \dots, x^{n+k})$ into $\mathbb{R}^n$,*

$$F = (F_1, \dots, F_n), \quad F_i\colon U \to \mathbb{R}, \quad i = 1, \dots, n.$$

*Let $F(x_0) = 0$, where $x_0 = (x_0^1, \dots, x_0^{n+k}) \in U$, and suppose that the matrix*

$$\left(\frac{\partial F_i}{\partial x^{k+j}}\right)_{1 \le i,j \le n}$$

*is invertible at the point $x_0$. Then there exist a neighborhood $V$ of the point $x_0$ and a neighborhood $W$ of the point $(x_0^1, \dots, x_0^k) \in \mathbb{R}^k$ such that:*

1) *there are smooth functions*

$$f_1, \dots, f_n\colon W \to \mathbb{R}$$

*defined on $W$;*

2) *if $x = (x^1, \dots, x^{n+k}) \in V$, then $(x^1, \dots, x^k) \in W$;*

3) *$F(x) = 0$ if and only if*

$$x^{k+1} = f_1(x^1, \dots, x^k), \ \dots, \ x^{k+n} = f_n(x^1, \dots, x^k).$$

The condition on the point $x_0$ means that it is nonsingular. Namely, a point $x \in U \subset \mathbb{R}^n$ is a *nonsingular* or a *regular* point of a smooth mapping $F: U \to \mathbb{R}^m$,

$$(x^1, \dots, x^n) \to (y^1(x^1, \dots, x^n), \dots, y^m(x^1, \dots, x^n)),$$

if at this point the rank of the matrix of the first-order derivatives $\left(\frac{\partial y^i}{\partial x^j}\right)$ equals $m$. Otherwise the point $x$ is said to be *singular* or *critical* point of the mapping $F$.

For instance, let $f: U \to \mathbb{R}$ be a smooth function. Then a point $x$ is nonsingular if the gradient of $f$ at this point does not vanish, $\operatorname{grad} f \neq 0$.

**Lemma 3.1.** *A set of points $S \subset \mathbb{R}^3$ is a regular surface if and only if for every point $x \in S$ there exists a neighborhood $U \subset \mathbb{R}^3$ of this point such that in this neighborhood the set $S$ is representable as the set of zeros of a smooth function $F: U \to \mathbb{R}$, and all the points of $S$ are nonsingular points of $F$.*

**Proof.** As we pointed out, if a surface in $U$ is the graph of a function $z = f(x, y)$, then we can take $F$ to be

$$F(x, y, z) = z - f(x, y).$$

Assume now that there is a function $F: U \to \mathbb{R}$, defined in a neighborhood of a point $(x_0, y_0, z_0)$, such that the set of its zeros $\{F = 0\}$ consists of nonsingular points. Without loss of generality we may assume that

$$\frac{\partial F(x_0, y_0, z_0)}{\partial z} \neq 0.$$

Then by the implicit function theorem there exists a function $f$ defined in a neighborhood of the point $(x_0, y_0)$ and a domain $U' \subset U$ such that $F(x, y, z) = 0$ for $(x, y, z) \in U'$ if and only if $z = f(x, y)$. Therefore, the set of zeros $\{F = 0\}$ in a neighborhood of every point of $S$ can be represented as the graph of a function. □

**Lemma 3.2.** *A set of points $S \subset \mathbb{R}^3$ is a regular surface if and only if, in a sufficiently small neighborhood $U \subset \mathbb{R}^3$ of every point, $S$ can be specified as the image of a smooth mapping*

$$r: (u, v) \to (x(u, v), y(u, v), z(u, v))$$

*from a domain $V \subset \mathbb{R}^2$ into $U$, and the vectors $r_u = dr/du$ and $r_v = dr/dv$ are linearly independent at every point of $V$.*

**Proof.** The graph of a function is a particular case of the parametric representation. Therefore, it suffices to show that for a mapping $r: V \to \mathbb{R}^3$ of the form

$$(u, v) \to (x(u, v), y(u, v), z(u, v))$$

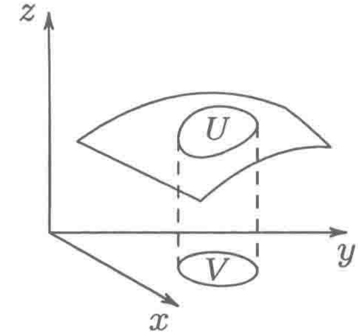


**Figure 3.1.** Surface as the graph of a function.

such that all the points $(u,v) \in V$ are nonsingular, the image of $V$, in a neighborhood of every point of this image, is the graph of a function.

Let $(u_0, v_0) \in V$. The vectors $r_u$ and $r_v$ are linearly independent everywhere. We may assume without loss of generality that the matrix $\left(\begin{smallmatrix} x_u & x_v \\ y_u & y_v \end{smallmatrix}\right)$ is invertible at the point $(u_0, v_0)$. By the inverse function theorem, there exists the inverse mapping

$$(x,y) \to (u(x,y), v(x,y))$$

defined in some neighborhood of the point $(x(u_0,v_0), y(u_0,v_0))$. Therefore, in a sufficiently small neighborhood of the point $r(u_0,v_0)$, the surface can be specified as the graph of the function

$$z = z(u(x,y), v(x,y)) = \tilde{z}(x,y),$$

which completes the proof. □

So far we have not introduced any scalar product in $\mathbb{R}^3$, but it is easily seen that in the definition of a surface one may consider the Euclidean coordinates $x$, $y$, $z$ if the space $\mathbb{R}^3$ is endowed with Euclidean scalar product.

By a domain on a regular surface $S$ we mean the intersection of $S$ with a domain in $\mathbb{R}^3$. Accordingly, for a point of $S$, any domain on $S$ containing this point is said to be a neighborhood of this point.

**3.1.2. Local coordinates.** Parametrization of a surface is particularly convenient in that it introduces *local coordinates.* Namely, if a mapping $r\colon V \to \mathbb{R}^3$ specifies a regular surface, then the coordinates $u$ and $v$ become coordinates in the domain $r(V)$ on the surface: the tuples $(u,v) \in V$ are in one-to-one correspondence with the points of the plane. The mapping $r\colon V \to S$ is called a *local map* covering the domain $r(V)$.

A function on the surface

$$f\colon S \to \mathbb{R}$$

is said to be *continuous* or *smooth* if, in a neighborhood of each point, it can be represented by a continuous or smooth function of local coordinates.

To be well defined, the properties of continuity and smoothness must be independent of the choice of coordinates. Namely, let two domains $r(V)$ and $\tilde{r}(\tilde{V})$ have a nonempty intersection. Then in this intersection we obtain two different systems of local coordinates, $(u,v)$ and $(\tilde{u},\tilde{v})$. By the inverse function theorem, there exist smooth one-to-one mappings

$$(u,v) \to (\tilde{u}(u,v), \tilde{v}(u,v)), \quad (\tilde{u},\tilde{v}) \to (u(\tilde{u},\tilde{v}), v(\tilde{u},\tilde{v}))$$

defined on $r^{-1}(r(V) \cap \tilde{r}(\tilde{V}))$ and $\tilde{r}^{-1}(r(V) \cap \tilde{r}(\tilde{V}))$. Therefore, if, say,

$$f(u,v) = f(u(\tilde{u},\tilde{v}), v(\tilde{u},\tilde{v})),$$

then

$$\frac{\partial f}{\partial \tilde{u}} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial \tilde{u}} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial \tilde{u}}, \qquad \frac{\partial f}{\partial \tilde{v}} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial \tilde{v}} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial \tilde{v}}.$$

Since the Jacobian

$$\begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial u}{\partial \tilde{v}} \\ \frac{\partial v}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix}$$

does not vanish and is also a smooth function, the notion of smoothness, and *a fortiori* of continuity, does not depend on the choice of coordinates. We may also speak about the degree of smoothness of a function, which makes sense, however, only when the smoothness of the function is less than that of the changes of local coordinates, since otherwise this property is no longer invariant. Usually we deal with infinitely smooth surfaces (of class $C^\infty$), in which case this question does not arise. However, analytic functions can be defined only under additional restrictions on the changes of coordinates.

A mapping from one surface to another, $F\colon S_1 \to S_2$, is said to be *smooth* if it is everywhere specified by smooth functions of local coordinates:

$$(x,y) \to (u(x,y), v(x,y)),$$

where $(x,y)$ are local coordinates on $S_1$, and $(u,v)$, on $S_2$. This notion is well defined for the same reasons as that of a smooth function on a surface. Continuity of a mapping between surfaces is defined in a similar way.

Note that we have discussed only regular surfaces so far. However, sometimes it is convenient to deal with more general surfaces, which are locally or globally specified by an equation $F(x,y,z) = 0$. In this case the points where $\operatorname{grad} F = 0$ are said to be *singular points* of the surface.

EXAMPLES. 1. Consider a sphere or, more generally, an ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

This is a regular surface which cannot be globally parametrized or globally represented as the graph of a function.

2. A HYPERBOLOID OF ONE SHEET,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1,$$

is not globally a graph of a function, but admits a global parametric representation

$$x = a\sqrt{1 + \frac{z^2}{c^2}} \cdot \cos\varphi, \quad y = b\sqrt{1 + \frac{z^2}{c^2}} \cdot \sin\varphi, \quad z = z,$$

where $z$ and the angle $\varphi$ (in the polar coordinates on the plane $xy$) specify coordinates on the entire hyperboloid.

3. A CONE,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0.$$

This surface has a single singular point $(0,0,0)$. If we delete it, the cone falls into two parts (with $z > 0$ and $z < 0$), which are the graphs of the functions

$$z = \pm c\sqrt{\frac{x^2}{a^2} + \frac{y^2}{b^2}}.$$

4. SURFACES OF REVOLUTION. Let $r(t)$ be a regular curve in the half-plane $y \geq 0$, $z = 0$. Revolving it about the $x$-axis yields a *surface of revolution*. For example, if this curve is the graph of the function $y = \sqrt{R^2 - x^2}$, $-R \leq x \leq R$, we obtain the sphere $x^2 + y^2 + z^2 = R^2$. In general, a surface of revolution obtained from a function $y = f(x)$ can be parametrized as

$$r(u, v) = (u, f(u)\cos v, f(u)\sin v).$$

A *torus of revolution* is obtained by rotation of a circle $(x - a)^2 + (y - b)^2 = R^2$ with $b > R$.

**3.1.3. Tangent space.** Consider a smooth curve on a surface. If the surface is parametrized, then the curve can be represented as the composition of mappings

$$t \to (u(t), v(t)) \to r(u(t), v(t)).$$

The velocity vector is

$$\frac{dr(u(t), v(t))}{dt} = r_u\dot{u} + r_v\dot{v}.$$

Furthermore, any vector of the form $\xi = \xi^1 r_u(u_0, v_0) + \xi^2 r_v(u_0, v_0)$ is the velocity vector of some curve on the surface. For example, we can take the curve written in local coordinates as

$$u = u_0 + \xi^1 t, \qquad v = v_0 + \xi^2 t.$$

The vectors $\xi$ as above form a two-dimensional vector space, which is called the *tangent space* at the point $r(u, v)$. The vectors $r_u$ and $r_v$ are a basis of this space.

EXAMPLES. 1. Let the surface be specified by the equation $F(x, y, z) = 0$. Then a curve $\gamma(t) = (x(t), y(t), z(t))$ on the surface satisfies the equation $f(t) = F(x(t), y(t), z(t)) = 0$, and so $df/dt \equiv 0$. Therefore, the tangent space at a point $(x_0, y_0, z_0)$ consists of the vectors $(\dot{x}, \dot{y}, \dot{z})$ such that

$$\frac{df}{dt} = \left\langle \operatorname{grad} F, \frac{d\gamma}{dt} \right\rangle = \frac{\partial F}{\partial x}\dot{x} + \frac{\partial F}{\partial y}\dot{y} + \frac{\partial F}{\partial z}\dot{z} = 0.$$

2. Let a surface be the graph of a function $z = f(x, y)$. Then the vectors $(1, 0, f_x)$ and $(0, 1, f_y)$ form a basis in the tangent space at a given point.

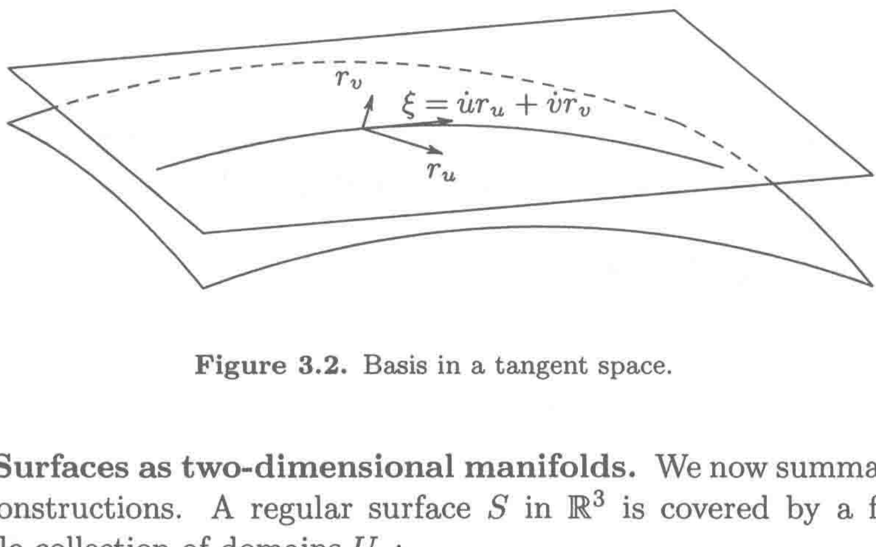


**Figure 3.2.** Basis in a tangent space.

**3.1.4. Surfaces as two-dimensional manifolds.** We now summarize the above constructions. A regular surface $S$ in $\mathbb{R}^3$ is covered by a finite or countable collection of domains $U_\alpha$:

$$S = \bigcup_\alpha U_\alpha$$

such that:

1) in each domain $U_\alpha$ we can define *local coordinates* $x^1_\alpha, x^2_\alpha$;

2) the local coordinates $(x^1_\alpha, x^2_\alpha)$ run over some domain $V_\alpha \subset \mathbb{R}^2$, and each point of $V_\alpha$ is associated with exactly one point of the domain $U_\alpha$ on the surface;

3) in the intersection of two domains $U_\alpha \cap U_\beta$ the local coordinates $(x^1_\alpha, x^2_\alpha)$ and $(x^1_\beta, x^2_\beta)$ are related by mutually inverse smooth mappings

(changes of coordinates):

$$x^i_\alpha = x^i_\alpha(x^1_\beta, x^2_\beta), \quad x^j_\beta = x^j_\beta(x^1_\alpha, x^2_\alpha), \quad i, j = 1, 2,$$

with nonzero Jacobians

$$\det\Big(\frac{\partial x^i_\alpha}{\partial x^j_\beta}\Big) \neq 0, \qquad \det\Big(\frac{\partial x^i_\beta}{\partial x^j_\alpha}\Big) \neq 0.$$

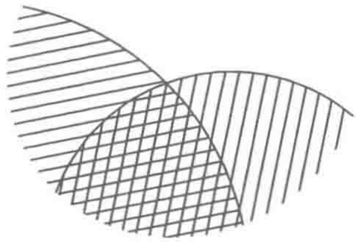

**Figure 3.3.** Overlapping charts.

The collection of domains $U_\alpha$ is called an *atlas of the surface*, and the domains $U_\alpha$ are called *charts*.

Now we can carry over many notions of the classical calculus to regular surfaces.

a) A *domain* on a surface $S$ is a set $U \subset S$ such that the coordinates $(x^1_\alpha, x^2_\alpha)$ of the points in the intersection of $U$ with any chart $U_\alpha$ form a domain in $\mathbb{R}^2$.

b) Any domain $U$ containing a point $x \in S$ is said to be a *neighborhood* of the point $x$.

c) A function $f\colon S \to \mathbb{R}$ is said to be *smooth* if in each chart $U_\alpha$ it may be written as a smooth function of local coordinates $x^1_\alpha, x^2_\alpha$.

d) A mapping $r\colon (a, b) \to S$ specifies a *continuous curve* if for every point $r(t_0)$ lying in a chart $U_\alpha$, the points of the interval $|t - t_0| < \varepsilon$ for some $\varepsilon > 0$ are carried by $r$ also into $U_\alpha$, and in local coordinates $x^1_\alpha, x^2_\alpha$ this mapping is given by continuous functions

$$r\colon (t_0 - \varepsilon, t_0 + \varepsilon) \to (x^1_\alpha(t), x^2_\alpha(t)).$$

Moreover, if $x^1_\alpha(t), x^2_\alpha(t)$ are smooth functions of $t$, then the curve $r(t)$ is said to be *smooth*.

Note that the regular surfaces in $\mathbb{R}^3$ additionally possess the following "Hausdorff" property.

4) For any two different points $x, y$ on the surface, there exist nonoverlapping neighborhoods $U$ and $V$,

$$U \cap V = \varnothing, \quad x \in U, \quad y \in V.$$

Any set of points for which there is an atlas satisfying conditions 1)–4) is called a *two-dimensional smooth manifold*.

A *tangent vector* $\xi$ at a point $x$ of a two-dimensional manifold is the velocity vector of a smooth curve $r(t)$ at the point $x$:

$$\xi = \frac{dr(t_0)}{dt}, \quad r(t_0) = x.$$

It is written differently in different local coordinates. If the point $x$ lies in the intersection of two charts $U_\alpha$ and $U_\beta$, we have

$$\xi_\alpha = (\dot{x}_\alpha^1, \dot{x}_\alpha^2)$$

in the coordinates $x_\alpha^1, x_\alpha^2$. Then by the chain rule (for differentiation of a composite function) the same tangent vector in the coordinates $x_\beta^1, x_\beta^2$ is written as

$$\xi_\beta = \Big(\frac{dx_\beta^1(x_\alpha^1(t), x_\alpha^2(t))}{dt}, \frac{dx_\beta^2(x_\alpha^1(t), x_\alpha^2(t))}{dt}\Big) = \Big(\frac{\partial x_\beta^1}{\partial x_\alpha^i}\dot{x}_\alpha^i, \frac{\partial x_\beta^2}{\partial x_\alpha^i}\dot{x}_\alpha^i\Big).$$

Therefore, a tangent vector at a point $x$ may be defined as an entity $\xi = (\xi^1, \xi^2)$ whose representations $\xi_\alpha$ and $\xi_\beta$ in different local coordinates are related as

$$\xi_\beta^i = \frac{\partial x_\beta^i}{\partial x_\alpha^j}(x)\xi_\alpha^j.$$

All tangent vectors of a two-dimensional manifold $S$ at a point $x$ form a vector space called the *tangent space* at the point $x$ and denoted by $T_xS$.

The above definitions are rather general, and, as can be shown, by no means every two-dimensional smooth manifold can be realized as a regular surface in $\mathbb{R}^3$. By realization we mean the existence of a smooth *embedding*, i.e., a smooth mapping

$$F \colon S \to \mathbb{R}^3$$

such that $F$ has maximal rank at each point,

$$\operatorname{rank}\Big(\frac{\partial x^i}{\partial x_\alpha^j}\Big) = 2;$$

different points of the surface are mapped into different points of the Euclidean space,

$$F(x) \neq F(y) \quad \text{if} \quad x \neq y;$$

and whenever a sequence of points $F(x_i)$ converges to a point $F(x_\infty)$, the sequence $x_i$ converges to the point $x_\infty$.

To be more precise, we say that a sequence of points $y_i$ converges to a point $y_\infty$ if the points $y_i$ lie in a coordinate neighborhood of $y_\infty$ for all $i$ large enough, and in this neighborhood they converge to $y_\infty$ in the usual sense (as on the plane).

If there exists such an embedding, its image is called a smooth *submanifold* of the Euclidean space $\mathbb{R}^3$ or a regular surface in $\mathbb{R}^3$.

## 3.2. Riemannian metric on a surface

**3.2.1. The length of a curve on a surface.** Let $x, y, z$ be the Euclidean coordinates in $\mathbb{R}^3$.

It is natural to define the length of a curve $(u(t), v(t))$ on a surface $r(u, v)$ as its length in $\mathbb{R}^3$. In this case the metric on the surface is said to be *induced* by the metric of the space $\mathbb{R}^3$ containing the surface.

The velocity vector is

$$(\dot{x}, \dot{y}, \dot{z}) = r_u \dot{u} + r_v \dot{v},$$

where

$$\dot{x} = x_u \dot{u} + x_v \dot{v}, \qquad \dot{y} = y_u \dot{u} + y_v \dot{v}, \qquad \dot{z} = z_u \dot{u} + z_v \dot{v},$$

and the length of the curve is

$$l = \int \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}\, dt.$$

Substituting the formulas for $\dot{x}$, $\dot{y}$, and $\dot{z}$ into the integrand, we obtain

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2,$$

where

$$\begin{aligned} E &= \langle r_u, r_u \rangle = x_u^2 + y_u^2 + z_u^2, \\ F &= \langle r_u, r_v \rangle = x_u x_v + y_u y_v + z_u z_v, \\ G &= \langle r_v, r_v \rangle = x_v^2 + y_v^2 + z_v^2. \end{aligned}$$

In order to present these formulas in a tensor form, we will write the coordinates $u$ and $v$ as $x^1$ and $x^2$. Then

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix},$$

where

$$g_{ij} = \left\langle \frac{\partial r}{\partial x^i}, \frac{\partial r}{\partial x^j} \right\rangle.$$

Thus we have proved the following lemma.

**Lemma 3.3.** *The length of a curve $(x^1(t), x^2(t))$ on a surface is equal to*

$$l = \int \sqrt{g_{ij}\dot{x}^i \dot{x}^j}\, dt.$$

The expression

$$dl^2 = g_{ij}\, dx^i\, dx^j = E\, du^2 + 2F\, du\, dv + G\, dv^2$$

is called the *first fundamental form* or the *Riemannian metric* on the surface. Here $E$, $F$, and $G$ are functions of the coordinates $u$ and $v$.

For each point of the surface, this form specifies a Euclidean scalar product on the tangent space at this point:

$$\xi = \xi^i \frac{\partial r}{\partial x^i}, \ \eta = \eta^j \frac{\partial r}{\partial x^j} \quad \to \quad \langle \xi, \eta \rangle = g_{ij} \xi^i \eta^j.$$

By means of this scalar product, the angle $\varphi$ between the tangent vectors $\xi$ and $\eta$ is defined by the usual formula:

$$\cos\varphi = \frac{\langle \xi, \eta \rangle}{|\xi||\eta|}.$$

EXAMPLES. 1. Let the surface be the graph of a function $z = f(x, y)$. Then

$$r_x = (1, 0, f_x), \quad r_y = (0, 1, f_y),$$
$$E = \langle r_x, r_x \rangle = 1 + f_x^2, \quad F = \langle r_x, r_y \rangle = f_x f_y, \quad G = \langle r_y, r_y \rangle = 1 + f_y^2.$$

2. Let the surface by given by the equation $F(x, y, z) = 0$ and let $F_z \neq 0$ in a neighborhood of a point $(x_0, y_0, z_0)$. Take $x$ and $y$ for the local coordinates, $u = x$, $v = y$. The condition $F = 0$ implies the identity

$$F_x \dot{x} + F_y \dot{y} + F_z \dot{z} = 0$$

for the tangent vectors $(\dot{x}, \dot{y}, \dot{z})$ to the surface, which implies

$$\dot{z} = -\frac{1}{F_z} (F_x \dot{x} + F_y \dot{y}).$$

Hence we derive that

$$\begin{aligned} \dot{x}^2 + \dot{y}^2 + \dot{z}^2 &= \dot{x}^2 + \dot{y}^2 + \frac{1}{F_z^2} (F_x^2 \dot{x}^2 + 2F_x F_y \dot{x}\dot{y} + F_y^2 \dot{y}^2) \\ &= \Big(1 + \frac{F_x^2}{F_z^2}\Big)\dot{x}^2 + 2\,\frac{F_x F_y}{F_z^2}\,\dot{x}\dot{y} + \Big(1 + \frac{F_y^2}{F_z^2}\Big)\dot{y}^2. \end{aligned}$$

Thus we obtain the following formulas for the metric:

$$g_{11} = 1 + \frac{F_x^2}{F_z^2}, \qquad g_{12} = \frac{F_x F_y}{F_z^2}, \qquad g_{22} = 1 + \frac{F_y^2}{F_z^2}.$$

Suppose that in some domain on the surface there are two different coordinate systems $(x^1, x^2)$ and $(\tilde{x}^1, \tilde{x}^2)$ related by the transition formulas.

Then the same tangent vector may be expanded in different bases:

$$\xi = \xi^i \frac{\partial r}{\partial x^i} = \tilde{\xi}^j \frac{\partial r}{\partial \tilde{x}^j}.$$

Since its length does not depend on the choice of the basis, we have

$$g_{ij}\xi^i\xi^j = \tilde{g}_{kl}\tilde{\xi}^k\tilde{\xi}^l.$$

This equality can be rewritten as $g_{ij}\,dx^i\,dx^j = \tilde{g}_{kl}\,d\tilde{x}^k\,d\tilde{x}^l$. Putting into the right-hand side of this equality the expressions $d\tilde{x}^k = \frac{\partial \tilde{x}^k}{\partial x^i}\,dx^i$, we obtain

$$g_{ij}\,dx^i\,dx^j = \tilde{g}_{kl}\frac{\partial \tilde{x}^k}{\partial x^i}\frac{\partial \tilde{x}^l}{\partial x^j}\,dx^i\,dx^j.$$

The equality between the forms implies the equality of the corresponding coefficients. Thus we have proved the following fact.

**Lemma 3.4.** *The coefficients of the first fundamental form*

$$g_{ij}\,dx^i\,dx^j = \tilde{g}_{kl}\,d\tilde{x}^k\,d\tilde{x}^l$$

*written in different coordinate systems are related by the formula*

$$g_{ij} = \tilde{g}_{kl}\frac{\partial \tilde{x}^k}{\partial x^i}\frac{\partial \tilde{x}^l}{\partial x^j}.$$

A smooth mapping of one surface to another is called an *isometry* if it preserves the lengths of all curves. This property can be written in local coordinates as follows. Let the mapping be

$$(x^1, x^2) \to (y^1(x^1, x^2), y^2(x^1, x^2)).$$

The isometry condition is equivalent to the equality

$$\tilde{g}_{ij}(y(x))\,dy^i\,dy^j = \tilde{g}_{ij}(y(x))\frac{\partial y^i}{\partial x^k}\frac{\partial y^j}{\partial x^l}\,dx^k\,dx^l = g_{kl}\,dx^k\,dx^l, \tag{3.1}$$

where $g_{kl}\,dx^k\,dx^l$ and $\tilde{g}_{ij}\,dy^i\,dy^j$ are the first fundamental forms of the surfaces.

Indeed, let $r(t) = (x^1(t), x^2(t))$ be a curve and let $\tilde{r}(t)$ be its image, $a \le t \le b$. They have the same length if

$$\begin{aligned}\int_a^b \sqrt{g_{kl}(r(t))\dot{x}^k\dot{x}^l}\,dt &= \int_a^b \sqrt{\tilde{g}_{ij}(\tilde{r}(t))\dot{y}^i\dot{y}^j}\,dt \\ &= \int_a^b \sqrt{\tilde{g}_{ij}(\tilde{r}(t))\,\frac{\partial y^i}{\partial x^k}\,\frac{\partial y^j}{\partial x^l}\,\dot{x}^k\dot{x}^l}\,dt.\end{aligned}$$

Under an isometry this equality holds for any curve $r(t)$, which is equivalent to relations (3.1).

**3.2.2. Surface area.** The *area* of a domain $U$ on a surface $r = r(u, v)$ is defined as

$$\sigma(U) = \iint_U \sqrt{g}\, du\, dv.$$

Here the domain $U$ is parametrized by coordinates $u$ and $v$, and

$$g = \det(g_{ij}) = g_{11}g_{22} - g_{12}^2 = EF - G^2.$$

We take this definition, and that of the length of a curve, as an axiom, but we will present some arguments to justify it.

Let $\tilde{e}_1$, $\tilde{e}_2$ be a basis in $\mathbb{R}^2$ and $\tilde{g}_{ij} = \langle \tilde{e}_i, \tilde{e}_j \rangle$. Determine the area of the parallelogram with sides $\tilde{e}_1$ and $\tilde{e}_2$. Let $e_1$, $e_2$ be an orthonormal basis such that

$$\tilde{e}_1 = ae_1, \qquad \tilde{e}_2 = be_1 + ce_2.$$

Then the area of the parallelogram is $|ac|$. It is easily seen that

$$a^2c^2 = \langle \tilde{e}_1, \tilde{e}_1 \rangle \langle \tilde{e}_2, \tilde{e}_2 \rangle - \langle \tilde{e}_1, \tilde{e}_2 \rangle^2 = \tilde{g}.$$

Let $U$ be a domain on a surface

$$z = f(x, y)$$

and let $V$ be its projection on the plane $xy$. The area of the parallelogram spanned by the tangent vectors $r_x$ and $r_y$ at a point $(x, y)$ is equal to $\sqrt{g(x, y)}$. Here $r(x, y) = (x, y, f(x, y))$.

The *tangent plane* to a surface at a point $r(x_0, y_0)$ is the plane in $\mathbb{R}^3$ which passes through the point $r(x_0, y_0)$ and consists of the tangent vectors to the surface. It is given by the equation

$$z = z_0 + f_x(x - x_0) + f_y(y - y_0),$$

where $z_0 = f(x_0, y_0)$ and the partial derivatives of $f$ are taken at the point $(x_0, y_0)$. Using the Taylor expansion of $f$, we can write the equation of the surface as

$$z = z_0 + f_x(x - x_0) + f_y(y - y_0) + O(\rho^2),$$

where $\rho = \sqrt{(x - x_0)^2 + (y - y_0)^2}$. Therefore, the tangent plane approximates the surface within $O(\rho^2)$ as $\rho \to 0$.

The area of a domain $V$ on the plane is defined as

$$\iint_V dx\, dy,$$

i.e., as the integral of the function $f(x, y) \equiv 1$ over the domain $V$. For an arbitrary function $f(x, y)$, the integral over $V$ is defined as the limit (provided it exists) of the sums

$$\sum_\alpha f(x_\alpha, y_\alpha)\Delta x_\alpha \Delta y_\alpha$$

over partitions of subdomains $V' \subset V$ into rectangles. Here $(x_\alpha, y_\alpha)$ is the center of a generic rectangle of the partition, the sides of the rectangles tend to zero, $\Delta x_\alpha, \Delta y_\alpha \to 0$, and in the limit, the domains $V'$ exhaust $V$. The parallelograms in the tangent planes at the points $(x_\alpha, y_\alpha)$ with sides $r_x \Delta x_\alpha$, $r_y \Delta y_\alpha$ approximate the surface within $(\Delta x_\alpha)^2 + (\Delta y_\alpha)^2$, and their areas are $\sqrt{g(x_\alpha, y_\alpha)}$. Hence it is natural to define the area of the domain $U$ on the surface as the integral

$$\iint_V \sqrt{g}\, dx\, dy.$$

EXAMPLES. 1. If a surface is the graph of a function $z = f(x, y)$, and $V$ is the projection of a surface domain $U$ on the plane $(x, y)$, then

$$\sigma(U) = \iint_V \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy.$$

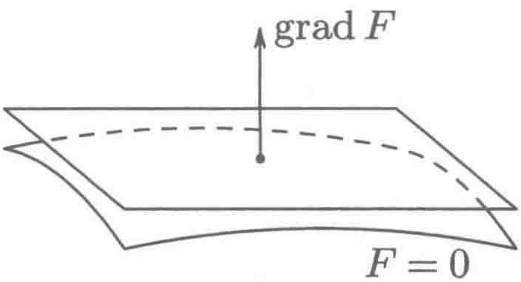


**Figure 3.4.** Gradient of a function and the tangent plane.

2. Let a surface be specified by the equation $F(x, y, z) = 0$, and let $F_z \neq 0$ in a domain $U$, which projects to a domain $V$ on the plane $(x, y)$. Then

$$\sigma(U) = \iint_V \frac{|\operatorname{grad} F|}{|F_z|}\, dx\, dy.$$

3. If a surface is given in a parametric form $r = r(u, v)$, and $V$ is the domain on the plane $(u, v)$ such that $r(V) = U$, then

$$\sigma(U) = \iint_V |[r_u, r_v]|\, du\, dv.$$

Here $[r_u, r_v]$ is the vector product in $\mathbb{R}^3$, which was defined so that the absolute value $|[\xi, \eta]|$ be equal to the area of the parallelogram spanned by the vectors $\xi$ and $\eta$.

## 3.3. Curvature of a surface

**3.3.1. On the notion of the surface curvature.** The idea of the curvature of a surface is most transparent in the case when the surface is the graph of a function and the point where we want to determine the curvature is a critical point of this function.

Consider the Euclidean coordinates in $\mathbb{R}^3$ such that the point $x = y = z = 0$ lies on the surface, the surface is the graph of a function $z = f(x, y)$ in a neighborhood of this point, and the $z$-axis is orthogonal to the surface at this point. The last condition means that the point $x = y = 0$ is a critical point of the function $f$, i.e., $f_x = f_y = 0$ for $x = y = 0$.

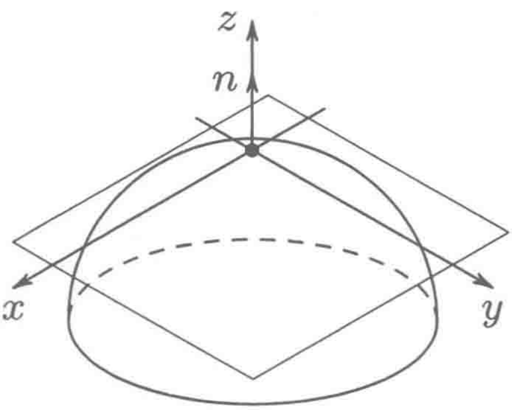


**Figure 3.5.** Special local coordinates.

At the point $x = y = 0$ the first fundamental form is $dx^2 + dy^2$ (i.e., $g_{ij} = \delta_{ij}$), and the *second fundamental form* is defined to be the quadratic term of the Taylor expansion of $f$:

$$b_{ij}\, dx^i\, dx^j = f_{xx}\, dx^2 + 2f_{xy}\, dx\, dy + f_{yy}\, dy^2$$

for $x = x^1 = 0, y = x^2 = 0$. The curvature is defined in terms of this form: the *Gaussian curvature* of the surface at the point $x = y = z = 0$ is the determinant,

$$K = f_{xx} f_{yy} - f_{xy}^2,$$

and the *mean curvature* is the half-trace,

$$H = \frac{1}{2}(f_{xx} + f_{yy}).$$

This definition gives a clear intuitive notion of the geometric meaning of positive or negative Gaussian curvature.

1) If $K > 0$, then the point $x = y = 0$ is either a (local) minimum point of the function $f(x, y)$ (for $H < 0$) or a (local) maximum point (for $H > 0$), and in a neighborhood of this point the surface lies entirely on one and the

same side of the tangent plane, i.e., is convex or concave (looking like a cup or a cap); see Figure 3.6.

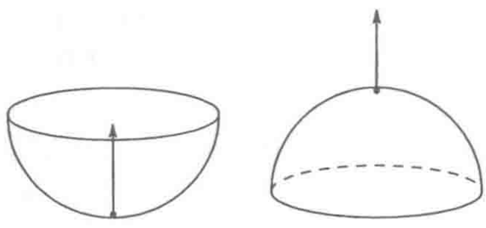

**Figure 3.6.** A cup ($k_1, k_2 < 0$) and a cap ($k_1, k_2 > 0$).

2) If $K < 0$, the point $x = y = 0$ is a saddle-point: in any arbitrarily small neighborhood of this point there are points of the surface lying on either side of the tangent plane, and the surface looks like a saddle (Figure 3.7).

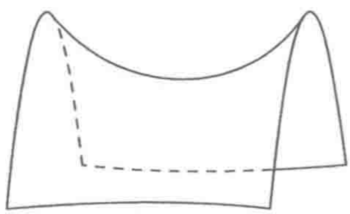

**Figure 3.7.** A saddle ($K < 0$).

To clarify how the curvature of a surface is related to the curvature of curves lying on this surface, we will invoke Euler's classical approach.

**3.3.2. Curvature of lines on a surface.** Consider a parametrized surface $r = r(u, v)$. At each point of the surface, we define the normal $n$ by the formula

$$n = \frac{[r_u, r_v]}{|[r_u, r_v]|}.$$

From this definition it follows that the vector $n$ is orthogonal to $r_u$ and $r_v$, $|n| = 1$, and the basis $(r_u, r_v, n)$ has positive orientation.

The *second fundamental form* is defined as

$$L\,du^2 + 2M\,du\,dv + N\,dv^2,$$

where

$$L = \langle r_{uu}, n\rangle, \qquad M = \langle r_{uv}, n\rangle, \qquad N = \langle r_{vv}, n\rangle.$$

Setting $x^1 = u$, $x^2 = v$, we will write it also in the form

$$b_{ij}\,dx^i\,dx^j, \quad \text{where} \quad L = b_{11}, \quad M = b_{12}, \quad N = b_{22}.$$

**Lemma 3.5.** *If $r = r(u(t), v(t))$ is a smooth curve on the surface, then the projection of the acceleration vector on the normal is equal to*

$$\langle \ddot{r}, n \rangle = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2,$$

*i.e., to the second fundamental form in the components of the velocity vector.*

**Proof.** The vector of acceleration equals

$$\ddot{r} = r_{uu}\dot{u}^2 + 2r_{uv}\dot{u}\dot{v} + r_{vv}\dot{v}^2 + r_u\ddot{u} + r_v\ddot{v}.$$

Taking the scalar product of $\ddot{r}$ and $n$, we obtain

$$\langle \ddot{r}, n \rangle = \langle r_{uu}, n \rangle \dot{u}^2 + 2\langle r_{uv}, n \rangle \dot{u}\dot{v} + \langle r_{vv}, n \rangle \dot{v}^2,$$

since $\langle r_u, n \rangle = \langle r_v, n \rangle = 0$ by definition. □

Let $l = l(t)$ be a natural parameter on the curve. Then by definition

$$\left| \frac{dr}{dl} \right| = \left| \frac{dr}{dt} \right| \frac{dt}{dl} \equiv 1.$$

The curvature $k$ of a curve $r$ in $\mathbb{R}^3$ and the principal normal to $r$ are defined by the equality

$$\frac{d^2 r}{dl^2} = k n_r,$$

where $|n_r| = 1$ (see Section 1.4.3). Lemma 3.5 implies that

$$\left\langle \frac{d^2 r}{dl^2}, n \right\rangle dl^2 = L\,du^2 + 2M\,du\,dv + N\,dv^2 = b_{ij}\,dx^i\,dx^j,$$

i.e.,

$$\left\langle \frac{d^2 r}{dl^2}, n \right\rangle = L\left(\frac{du}{dl}\right)^2 + 2M\,\frac{du}{dl}\,\frac{dv}{dl} + N\left(\frac{dv}{dl}\right)^2.$$

But

$$dl^2 = |\dot{r}|^2\,dt^2,$$

and we conclude that

$$k\langle n_r, n \rangle = \left\langle \frac{d^2 r}{dl^2}, n \right\rangle = \frac{L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2}{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} = \frac{b_{jk}\dot{x}^j\dot{x}^k}{g_{jk}\dot{x}^j\dot{x}^k},$$

where

$$\begin{gathered} g_{11} = E, \qquad g_{12} = F, \qquad g_{22} = G, \\ b_{11} = L, \qquad b_{12} = M, \qquad b_{22} = N, \\ \dot{x}^1 = \dot{u}, \qquad \dot{x}^2 = \dot{v}. \end{gathered}$$

The product $\langle n_r, n \rangle$ is the cosine of the angle between $n_r$ and $n$:

$$\langle n_r, n \rangle = \cos\theta.$$

Thus we have proved the following theorem.

**Theorem 3.2.** *If a curve lies on a surface in $\mathbb{R}^3$, then the product of its curvature and the cosine of the angle between the normal to the surface and the principal normal to the curve is equal to the ratio of the second and the first fundamental forms in the components of the velocity vector of the curve.*

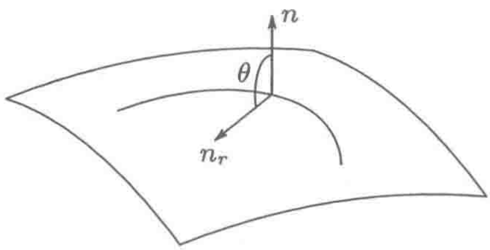


**Figure 3.8.** The normal to a curve and the normal to the surface.

Let $r(u_0, v_0)$ be a point on the surface. Consider the plane passing through this point with the space of its vectors spanned by the normal $n$ and a tangent vector $\xi = \xi^1 r_u + \xi^2 r_v$. This plane intersects the surface in a curve in $\mathbb{R}^3$, which is called a *normal section*. For this curve $\cos\theta = \pm 1$, where the sign depends on the side of the tangent plane at the point $r(u_0, v_0)$ where the curve lies. The quantity

$$k = \frac{b_{ij}\xi^i\xi^j}{g_{ij}\xi^i\xi^j}$$

is called the normal section curvature at the point $r(u_0, v_0)$, and its absolute value is equal to the curvature of this section as a curve in $\mathbb{R}^3$ (recall that the curvature of a curve in $\mathbb{R}^3$ is nonnegative by definition).

**3.3.3. Eigenvalues of a pair of scalar products.** Consider a general setting where two scalar products $G = (g_{ij})$ and $B = (b_{ij})$ on a vector space are given and $G$ is nondegenerate (the determinant of this matrix does not vanish). We say that the scalar product $G$ determines the geometry of the space (Euclidean, pseudo-Euclidean, symplectic, etc). For our present purpose, i.e., for the theory of surfaces in Euclidean space, we only need the case where both scalar products are symmetric and $g_{ij}\xi^i\xi^j > 0$ for $\xi \neq 0$. However, other cases are also very important for applications.

**Lemma 3.6.** *When changing the basis from $(e_1, \dots, e_n)$ to $(\tilde{e}_1, \dots, \tilde{e}_n)$ the Gram matrix of the scalar product, $Q = (\langle e_i, e_j \rangle)$, transforms by the formula*

$$Q \to \widetilde{Q} = A^\top Q A,$$

*where $Ae_i = \tilde{e}_i$ for $i = 1, \dots, n$.*

**Proof.** Consider the expansions of an arbitrary vector in the two bases:

$$\xi = \xi^i e_i = \tilde{\xi}^j \tilde{e}_j.$$

Since $\tilde{e}_j = a^i_j e_i$, we obtain

$$\xi^i e_i = \tilde{\xi}^j a^i_j e_i \quad \text{and} \quad \xi^i = a^i_j \tilde{\xi}^j.$$

The scalar product can be written in the matrix form as

$$\langle \xi, \eta \rangle = (\xi^1, \dots, \xi^n) Q \begin{pmatrix} \eta^1 \\ \vdots \\ \eta^n \end{pmatrix} = (\tilde{\xi}^1, \dots, \tilde{\xi}^n)(a^j_i) Q (a^i_j) \begin{pmatrix} \tilde{\eta}^1 \\ \vdots \\ \tilde{\eta}^n \end{pmatrix} = \tilde{\xi}^\top A^\top Q A \tilde{\eta},$$

where $A = (a^i_j)$ and the vectors are written as $(n \times 1)$ matrices (columns).

Therefore,

$$\widetilde{Q} = A^\top Q A,$$

which proves the lemma. □

Let $g^{ij}$ be the inverse matrix to $g_{ij}$, i.e., $g^{ik} g_{kj} = \delta^i_j$. Consider the operator determined by the matrix $b^i_j = g^{ik} b_{kj}$. Its eigenvalues are called the spectrum of the pair of scalar products specified by the matrices $G = (g_{ij})$ and $B = (b_{ij})$. In other words, we arrive at the following definition: the roots of the equation

$$\det(B - \lambda G) = 0$$

are called the *eigenvalues of the pair of scalar products.*

**Lemma 3.7.** *The eigenvalues of a pair of scalar products do not depend on the choice of the basis of the vector space.*

**Proof.** In another basis the Gram matrices have the form $A^\top B A = \widetilde{B}$ and $A^\top G A = \widetilde{G}$, hence

$$\begin{aligned} \det(\widetilde{B} - \lambda \widetilde{G}) &= \det(A^\top B A - \lambda A^\top G A) \\ &= \det A^\top \det(B - \lambda G) \det A = \det{}^2 A \det(B - \lambda G). \end{aligned}$$

Since $\det A \neq 0$, the polynomials $\det(B - \lambda G)$ and $\det(\tilde{B} - \lambda \tilde{G})$ have the same roots. □

**Lemma 3.8.** *Let $G$ and $B$ be two quadratic forms, i.e., the corresponding matrices are symmetric, and let the form $G$ be positive definite. Then there exists a basis of the vector space in which $G = (g_{ij})$ is the identity matrix and $B = (b_{ij})$ is a diagonal matrix.*

**Proof.** We regard $G$ as a scalar product, $\langle \xi, \eta \rangle = g_{ij}\xi^i\eta^j$. Then there exists an orthonormal basis in which $G$ is specified by the identity matrix. The form $B$ in this basis is given by a symmetric matrix. It is well known from a course of linear algebra that any quadratic form can be reduced by a rotation to a diagonal form. Hence the lemma. □

For the two-dimensional space we will construct this basis in an explicit form. Let $B = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$. The eigenvectors $\xi_1$, $\xi_2$ satisfy the equations $B\xi_i = \lambda_i\xi_i$, where $\lambda_1$, $\lambda_2$ are the roots of the equation

$$\det(B - \lambda \cdot 1) = 0.$$

It can be easily shown that these roots are real and equal to

$$\lambda_{1,2} = \frac{a + c \pm \sqrt{(a-c)^2 + 4b^2}}{2}.$$

If $\lambda_1 = \lambda_2$, then $a = c$, $b = 0$, so $B$ is already the identity matrix up to a scalar factor. If $\lambda_1 \neq \lambda_2$, then

$$\langle B\xi_1, \xi_2 \rangle = \lambda_1 \langle \xi_1, \xi_2 \rangle,$$

and, since the form is symmetric, we have

$$\lambda_1 \langle \xi_1, \xi_2 \rangle = \langle B\xi_1, \xi_2 \rangle = \langle \xi_1, B\xi_2 \rangle = \lambda_2 \langle \xi_1, \xi_2 \rangle.$$

The equality $(\lambda_1 - \lambda_2)\langle \xi_1, \xi_2 \rangle = 0$ implies that $\langle \xi_1, \xi_2 \rangle = 0$. Therefore, the vectors $\xi_1$ and $\xi_2$ are orthogonal, and in the basis $e_1 = \xi_1/|\xi_1|$, $e_2 = \xi_2/|\xi_2|$ the form $G$ is specified by the identity matrix, while $B$, by a diagonal matrix.

Obviously, in the basis stated in Lemma 3.8 the following equality holds:

$$B = \begin{pmatrix} \lambda_1 & & 0 \\ & & \\ 0 & & \lambda_n \end{pmatrix},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the two forms. The directions of the vectors $e_1, \ldots, e_n$ are called the *principal directions.*

The following lemma has been just proved in the two-dimensional case.

**Lemma 3.9.** *If the eigenvalues of a pair of quadratic forms are different, then the corresponding principal directions are orthogonal (relative to the form $G$).*

**Proof.** Let $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1 \langle \xi_1, \xi_2 \rangle = \langle B\xi_1, \xi_2 \rangle = \langle \xi_1, B\xi_2 \rangle = \lambda_2 \langle \xi_1, \xi_2 \rangle,$$

hence $\langle \xi_1, \xi_2 \rangle = 0$. □

**3.3.4. Principal curvatures and the Gaussian curvature.** Now we return to our geometric setting, where $G$ and $B$ are the first and the second fundamental forms of a surface at a point $r(u_0, v_0)$. Then the previous algebraic lemmas imply the following result.

**Theorem 3.3.** *For any point on a surface there exist coordinates $x^1$ and $x^2$ in a neighborhood of this point such that the first and the second fundamental forms at this point become*

$$g_{ij}\,dx^i\,dx^j = (dx^1)^2 + (dx^2)^2,$$
$$b_{ij}\,dx^i\,dx^j = k_1(dx^1)^2 + k_2(dx^2)^2.$$

In fact, we constructed such coordinates already in Section 3.3.1, before starting the systematic theory, by specifying the surface as the graph of a function $z = f(x, y)$, where $f_x = f_y = 0$ at the point of interest $x = x^1$, $y = x^2$.

Note that even the first fundamental form alone cannot in general be reduced to such a form in an entire neighborhood.

The eigenvalues $k_1$ and $k_2$ are called the *principal curvatures of the surface* at a given point. Their product

$$K = k_1 k_2$$

is called the *Gaussian curvature*, and their half-sum,

$$H = \frac{k_1 + k_2}{2},$$

is called the *mean curvature* of the surface.

Consider the equation $\det(B - \lambda G) = 0$, whose roots are $k_1$ and $k_2$:

$$\det(B - \lambda G) = \lambda^2 \det G - \lambda(g_{11}b_{22} + g_{22}b_{11} - g_{12}b_{21} - g_{21}b_{12}) + \det B = 0.$$

Applying the Vieta theorem we obtain the following assertion.

**Theorem 3.4.** *The Gaussian curvature is equal to the ratio of the first and the second fundamental forms:*

$$K = \frac{\det B}{\det G} = \frac{b_{11}b_{22} - b_{12}^2}{g_{11}g_{22} - g_{12}^2},$$

*and the mean curvature is*

$$H = \frac{1}{2}\,\frac{g_{11}b_{22} + g_{22}b_{11} - g_{12}b_{21} - g_{21}b_{12}}{g_{11}g_{22} - g_{12}^2}.$$

If $\xi = \xi^i r_{x^i}$ is a tangent vector at some point, then the curvature of the corresponding normal section equals

$$k = \frac{k_1(\xi^1)^2 + k_2(\xi^2)^2}{(\xi^1)^2 + (\xi^2)^2}.$$

Denote by $\alpha$ the angle between $\xi$ and the direction of the vector $\partial r/\partial x^1$. Then

$$\cos^2\alpha = \frac{(\xi^1)^2}{(\xi^1)^2 + (\xi^2)^2}, \qquad \sin^2\alpha = \frac{(\xi^2)^2}{(\xi^1)^2 + (\xi^2)^2}.$$

Thus we have proved the following *Euler's formula.*

**Theorem 3.5.** *The curvature of the normal section determined by a tangent vector $\xi$ is equal to*

$$k = k_1 \cos^2\alpha + k_2 \sin^2\alpha,$$

*where $k_1$, $k_2$ are the principal curvatures and $\alpha$ is the angle between $\xi$ and the principal direction corresponding to the vector $\frac{\partial r}{\partial x^1}$.*

For definiteness, let $k_1 \le k_2$. Then $k_1$ and $k_2$ are the minimal and maximal curvatures of normal sections.

EXAMPLE. Let the surface be the graph of a function $z = f(x, y)$. Then

$$r_x = (1, 0, f_x), \qquad r_y = (0, 1, f_y), \qquad [r_x, r_y] = (-f_x, -f_y, 1),$$
$$r_{xx} = (0, 0, f_{xx}), \qquad r_{xy} = (0, 0, f_{xy}), \qquad r_{yy} = (0, 0, f_{yy}),$$
$$n = \frac{[r_x, r_y]}{|[r_x, r_y]|} = \frac{(-f_x, -f_y, 1)}{\sqrt{1 + f_x^2 + f_y^2}},$$
$$L = b_{11} = \frac{f_{xx}}{\sqrt{1 + f_x^2 + f_y^2}}, \qquad M = b_{12} = b_{21} = \frac{f_{xy}}{\sqrt{1 + f_x^2 + f_y^2}},$$
$$N = b_{22} = \frac{f_{yy}}{\sqrt{1 + f_x^2 + f_y^2}}.$$

Since

$$g = g_{11}g_{22} - g_{12}^2 = (1 + f_x^2)(1 + f_y^2) - f_x^2 f_y^2 = 1 + f_x^2 + f_y^2,$$

we obtain the following result.

**Corollary 3.1.** *If the surface is the graph of a function $z = f(x, y)$, then its Gaussian curvature is*

$$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}. \tag{3.2}$$

We have already stated this formula in Section 3.3.1 for the case $f_x = f_y = 0$.

## 3.4. Basic equations of the theory of surfaces

**3.4.1. Derivational equations as the "zero curvature" condition. Gauge fields.** For any point of the surface $r = r(x^1, x^2)$, the vectors

$$r_1 = \frac{\partial r}{\partial x^1}, \quad r_2 = \frac{\partial r}{\partial x^2}, \quad n$$

form a basis of the three-dimensional space. Consider the expansions of the partial derivatives $r_{jk} = \frac{\partial^2 r}{\partial x^j \partial x^k}$ in this basis:

$$\begin{aligned} r_{11} &= \Gamma^1_{11} r_1 + \Gamma^2_{11} r_2 + b_{11} n, \\ r_{12} = r_{21} &= \Gamma^1_{12} r_1 + \Gamma^2_{12} r_2 + b_{12} n, \\ r_{22} &= \Gamma^1_{22} r_1 + \Gamma^2_{22} r_2 + b_{22} n. \end{aligned}$$

These equations can be written briefly as

$$r_{jk} = \Gamma^i_{jk} r_i + b_{jk} n.$$

Since the vectors $r_1$ and $r_2$ are orthogonal to $n$, we obtain

$$b_{ij} = \langle r_{ij}, n \rangle,$$

so these are the coefficients of the second fundamental form.

The quantities

$$\Gamma^i_{jk} = \Gamma^i_{kj}$$

are called the *Christoffel symbols*.

By the definition of the vector $n$ we have

$$\langle n, n \rangle \equiv 1, \qquad \langle n, r_1 \rangle = \langle n, r_2 \rangle \equiv 0,$$

hence

$$\frac{\partial}{\partial x^i} \langle n, n \rangle = 0, \qquad \frac{\partial}{\partial x^i} \langle n, r_j \rangle = 0$$

for all $i, j = 1, 2$. Similarly to deriving the Frenet equations, we obtain from the first equality that

$$\left\langle \frac{\partial n}{\partial x^i}, n \right\rangle = 0;$$

therefore, $\partial n / \partial x^i$ is a linear combination of the vectors $r_1$ and $r_2$:

$$\frac{\partial n}{\partial x^i} = a^j_i r_j.$$

In order to find the coefficients $a^j_i$, we will use the other equalities:

$$\frac{\partial}{\partial x^i} \langle n, r_j \rangle = \left\langle \frac{\partial n}{\partial x^i}, r_j \right\rangle + \langle n, r_{ij} \rangle = 0,$$

which can be rewritten as

$$a^k_i \langle r_k, r_j \rangle + b_{ij} = a^k_i g_{kj} + b_{ij} = 0. \tag{3.3}$$

The Gram matrix $(g_{ij})$ is invertible since the scalar product is nondegenerate. The inverse to the Gram matrix is denoted by $(g^{ij})$:

$$g^{ik}g_{kj} = \delta^i_j.$$

On multiplying the left-hand side of (3.3) by $(g^{lj})$ and summing over $j$ we obtain

$$g^{jl}a_i^k g_{kj} + g^{jl}b_{ij} = a_i^l + g^{jl}b_{ij} = 0.$$

These equations yield the coefficients $(a_i^j)$:

$$a_i^j = -b_{ik}g^{kj}.$$

The systems of equations

$$\frac{\partial}{\partial x^i}\begin{pmatrix} r_1 \\ r_2 \\ n \end{pmatrix} = A_i \begin{pmatrix} r_1 \\ r_2 \\ n \end{pmatrix},$$

where

$$A_1 = \begin{pmatrix} \Gamma^1_{11} & \Gamma^2_{11} & b_{11} \\ \Gamma^1_{12} & \Gamma^2_{12} & b_{12} \\ -b_{1k}g^{k1} & -b_{1k}g^{k2} & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} \Gamma^1_{21} & \Gamma^2_{21} & b_{21} \\ \Gamma^1_{22} & \Gamma^2_{22} & b_{22} \\ -b_{2k}g^{k1} & -b_{2k}g^{k2} & 0 \end{pmatrix},$$

are called the *derivational equations*. According to their geometric meaning, they must be compatible:

$$\frac{\partial}{\partial x^1}\frac{\partial}{\partial x^2}\begin{pmatrix} r_1 \\ r_2 \\ n \end{pmatrix} = \frac{\partial}{\partial x^2}\frac{\partial}{\partial x^1}\begin{pmatrix} r_1 \\ r_2 \\ n \end{pmatrix},$$

which is equivalent to the equations

$$\left[\frac{\partial}{\partial x^1}A_2 - \frac{\partial}{\partial x^2}A_1 - (A_1A_2 - A_2A_1)\right]\begin{pmatrix} r_1 \\ r_2 \\ n \end{pmatrix} = 0.$$

Since $r_1$, $r_2$, $n$ form a basis, we obtain

$$\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2} = [A_1, A_2], \tag{3.4}$$

where $[A_1, A_2]$ is the commutator of the matrices $A_1$ and $A_2$, i.e., $[A_1, A_2] = A_1A_2 - A_2A_1$.

We will call (3.4) the *zero curvature equations*; in the theory of surfaces they are called the *Codazzi equations*. It is worth noting here that in this case we are faced with the following very general situation.

Suppose we are given a family of square matrices $A_i(x^1, \dots, x^k)$, $i = 1, \dots, k$, called the *gauge field*, and we are to find a family of matrices

$\psi(x^1, \dots, x^k)$ satisfying the equations

$$\frac{\partial \psi}{\partial x^i} = A_i \psi, \quad i = 1, \dots, k.$$

In our case $k = 2$ and $\psi$ is the transition matrix from the initial orthonormal basis in $\mathbb{R}^3$ to the basis $r_1, r_2, n$ depending on the point on the surface.

It follows from commutativity of mixed derivatives that the problem of finding such a family of matrices $\psi$ has a solution if and only if the zero curvature equations hold.

Under a change $\psi(x) \to Q(x)\psi(x) = \varphi(x)$ the new matrix satisfies the system of equations

$$\frac{\partial \varphi}{\partial x^i} = \widetilde{A}_i \varphi(x),$$

where the new component of the gauge field $\widetilde{A}_i$ is related to the old component by the *gauge transformation* of the form

$$A_i = -Q^{-1}\frac{\partial Q}{\partial x^i} + Q^{-1}\widetilde{A}_i Q(x).$$

The family of matrices

$$R_{ij} = \frac{\partial A_i}{\partial x^j} - \frac{\partial A_j}{\partial x^i} - [A_i, A_j]$$

indexed by $i, j = 1, \dots, k$ is called the *curvature of the gauge field.* Under gauge transformations, the curvature transforms by the rule

$$R_{ij} = Q^{-1}\widetilde{R}_{ij} Q(x),$$

as can be shown by a simple calculation.

One can observe that the components of the curvature treated as operators are representable as commutators of the *covariant derivatives*,

$$[\nabla_i, \nabla_j] = R_{ij}, \qquad \nabla_i = \partial_i - A_i.$$

We will return to the gauge fields later. They are of fundamental importance in modern mathematics and physics.

The following lemma shows that the coefficients of the matrices $A_1$ and $A_2$ are completely determined by the first and the second fundamental forms.

**Lemma 3.10.** *The following equalities hold:*

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl} \left( \frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \right).$$

**Proof.** It follows from the definition of the Christoffel symbols that

$$\frac{\partial g_{ij}}{\partial x^k} = \Big\langle \frac{\partial r_i}{\partial x^k}, r_j \Big\rangle + \Big\langle r_i, \frac{\partial r_j}{\partial x^k} \Big\rangle = \Gamma^l_{ik} g_{lj} + \Gamma^l_{jk} g_{li}.$$

Hence we obtain

$$\frac{1}{2}\Big( \frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \Big) = \Gamma^s_{ij} g_{sl}.$$

To complete the proof, it suffices to multiply both sides of this equality by $g^{kl}$ and sum over $l$. □

**3.4.2. The Codazzi and sine-Gordon equations.** The Codazzi equations are the necessary and sufficient conditions for the form $(b_{ij})$ to be the second fundamental form of a surface with metric $(g_{ij})$. If they are fulfilled in a neighborhood $U$ of a point $(x_0^1, x_0^2)$, then the derivational equations with initial conditions at $(x_0^1, x_0^2)$

$$\langle r_i, r_j \rangle = g_{ij}(x_0^1, x_0^2), \qquad \langle r_i, n \rangle = 0, \qquad \langle n, n \rangle = 1$$

can be shown to have a solution. The last two relations are included into initial conditions, but it is seen from the form of the equations that they will be satisfied everywhere. Then the surface is locally constructed by the formula

$$r(x^1, x^2) = r(x_0^1, x_0^2) + \int_{x_0}^{x} (r_1\, dy^1 + r_2\, dy^2),$$

where the integration is along any path between the points $(x_0^1, x_0^2)$ and $(x^1, x^2)$ in a sufficiently small neighborhood $V$ of $(x_0^1, x_0^2)$ (actually it suffices to require that the neighborhood $V$ be simply connected; we will discuss this in detail in Section 4.5.1 when dealing with minimal surfaces). Similarly to the proof of Theorem 1.10, it can be shown that the $(g_{ij})$ and $(b_{ij})$ determine the surface up to motions of $\mathbb{R}^3$.

The following lemma is proved by direct calculations.

**Lemma 3.11.** *The Codazzi equations have the form*

$$\frac{\partial}{\partial x^k}\Gamma^l_{ij} - \frac{\partial}{\partial x^j}\Gamma^l_{ik} + \Gamma^s_{ij}\Gamma^l_{ks} - \Gamma^s_{ik}\Gamma^l_{js} = b_{ij}b^l_k - b_{ik}b^l_j, \tag{3.5}$$

$$\frac{\partial b_{ij}}{\partial x^k} - \frac{\partial b_{ik}}{\partial x^j} = \Gamma^s_{ik} b_{js} - \Gamma^s_{ij} b_{ks}, \tag{3.6}$$

*where* $b^j_i = b_{ik} g^{jk}$.

The equations (3.5) are called the *Gauss equations.*

The left- and right-hand sides of (3.5) and (3.6) change signs when $j$ and $k$ interchange. In fact, these equations may be reduced to three independent equations. One of them is obtained as follows. Multiply both sides of (3.5) by $g_{ml}$ and sum over $l$ to obtain an equivalent system of equations indexed by $i$, $j$, $k$, $m$. It can be shown that their left- and right-hand

sides change signs when $i$ and $m$ interchange. Therefore, equations (3.5) are equivalent in fact to a single equation, which is obtained, e.g., for $i = j = 1$, $k = m = 2$. The two other equations are obtained by putting $i = j = 1$, $k = 2$ into (3.6).

EXAMPLE. SURFACES OF CONSTANT NEGATIVE CURVATURE $K \equiv -1$. From a course of analytic geometry we know that a quadratic form $Q$ in two variables with $\det Q < 0$ factorizes into the product of linear forms. Therefore, the second fundamental form of a surface of negative curvature ($K < 0$) can be written as

$$b_{ij}\,dx^i\,dx^j = (A\,dx^1 + B\,dx^2)(C\,dx^1 + D\,dx^2),$$

where the linear forms can be chosen to be smooth in a small neighborhood of a given point. The direction of the tangent vector $\xi = (\xi^1, \xi^2)$ is said to be *asymptotic* (at the point) if $b_{ij}\xi^i\xi^j = 0$. In our setting these directions are found from the equations

$$A\xi^1 + B\xi^2 = 0, \qquad C\xi^1 + D\xi^2 = 0.$$

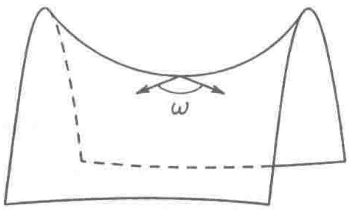


**Figure 3.9.** Asymptotic directions.

Their solutions can be taken to be the directions of the vectors $(B, -A)$ and $(D, -C)$. Since

$$\det\begin{pmatrix} B & -A \\ D & -C \end{pmatrix} = \det\begin{pmatrix} A & B \\ C & D \end{pmatrix} \neq 0,$$

in a neighborhood of any point one can find coordinates $u$ and $v$ such that $(B, -A)$ and $(D, -C)$ are the tangent vectors to the curves $u = \text{const}$ and $v = \text{const}$. In these coordinates

$$b_{ij}\,dx^i\,dx^j = 2b\,du\,dv.$$

Let

$$K = \frac{LN - M^2}{EG - F^2} = \frac{-b^2}{EG - F^2} \equiv -1.$$

Then $b^2 = EG - F^2$, and the equations (3.6) imply that

$$\frac{\partial E}{\partial v} = \frac{\partial G}{\partial u} = 0,$$

where $E$, $F$, and $G$ are the coefficients of the first fundamental form. Now we change the coordinates by setting

$$x = \int_{u_0}^{u} \sqrt{E}\, du, \qquad y = \int_{v_0}^{v} \sqrt{G}\, dv.$$

In these coordinates

$$(3.7) \quad g_{ij}\, dx^i\, dx^j = dx^2 + 2\cos\omega\, dx\, dy + dy^2, \quad b_{ij}\, dx^i\, dx^j = 2\sin\omega\, dx\, dy,$$

where $\omega$ is the angle between the asymptotic directions. Equations (3.5) for this metric reduce to the equation

$$\omega_{xy} + \sin\omega = 0,$$

which is referred to as the *sine-Gordon equation*. Introducing the variables $\xi = x + y$, $\tau = x - y$, it becomes

$$\frac{\partial^2\omega}{\partial\tau^2} - \frac{\partial^2\omega}{\partial\xi^2} = \sin\omega.$$

The sine-Gordon equation is one of the exactly integrable nonlinear equations, which are called *soliton* or KdV-like equations (i.e., similar to the Korteweg–de Vries (KdV) equation). Any nonzero solution of this equation determines the first and the second fundamental forms (3.7), which enable us to construct a surface of constant curvature $K \equiv -1$. The fact that the sine-Gordon equation does not have everywhere-geometric ($0 < \omega < \pi$) solutions played a fundamental role in Hilbert's theorem on nonexistence of globally regular embeddings of the Lobachevsky plane into $\mathbb{R}^3$ (see Section 4.3.4). In the 19th century Bianchi and Lie found substitutions (named "Bäcklund transformations" later on) that allowed for obtaining new solutions of this equation from those given in advance. The presence of such transformations is a characteristic feature of the famous integrable systems (the Korteweg–de Vries equation, Schrödinger's nonlinear equation (NLS), the sine-Gordon (SG) equation, the Kadomtsev–Petviashvili equation, etc.) and enables one to construct their simplest solutions. General rapidly decreasing (soliton-like) solutions are constructed by the methods of the *inverse scattering problem*, while solutions periodic in $x$ are obtained by the methods of analysis on Riemann surfaces (*finite-zone integration*).

**3.4.3. The Gauss theorem.** We set $i = j = 1$, $k = 2$ in equations (3.5), multiply both sides by $g_{2l}$, and sum over $l$. Then in the right-hand side we obtain the determinant of the second fundamental form and in the left-hand side, an expression involving the $g_{ij}$ and the Christoffel symbols, which also are representable in terms of the metric. Putting this expression into the formula for the Gaussian curvature, we obtain the Gauss theorem, which says that the Gaussian curvature can be expressed in terms of the coefficients of the first fundamental form and their derivatives.

However, this can be proved by a direct calculation. It follows from the derivational equations that

$$\langle r_{11}, r_{22}\rangle - \langle r_{12}, r_{12}\rangle = b_{11}b_{22} - b_{12}^2 + \Gamma_{11}^k\Gamma_{22}^l g_{kl} - \Gamma_{12}^k\Gamma_{12}^l g_{kl}.$$

At the same time,

$$\langle r_{11}, r_{22}\rangle - \langle r_{12}, r_{12}\rangle = \frac{\partial^2 g_{12}}{\partial x^1 \partial x^2} - \frac{1}{2}\frac{\partial^2 g_{11}}{\partial x^2 \partial x^2} - \frac{1}{2}\frac{\partial^2 g_{22}}{\partial x^1 \partial x^1},$$

which follows from the relation

$$\frac{\partial^2 g_{ij}}{\partial x^k \partial x^l} = \frac{\partial^2 \langle r_i, r_j\rangle}{\partial x^k \partial x^l} = (r_{ik}, r_{jl}) + (r_{il}, r_{jk}) + (r_i, r_{jkl}) + (r_{ikl}, r_j).$$

It remains to compare the two formulas for $\langle r_{11}, r_{22}\rangle - \langle r_{12}, r_{12}\rangle$ to complete the proof of the following theorem.

**Theorem 3.6.** *The following equality holds:*

$$K = \frac{b_{11}b_{22} - b_{12}^2}{g_{11}g_{22} - g_{12}^2} = \frac{g_{2l}}{g_{11}g_{22} - g_{12}^2}\left(\frac{\partial \Gamma_{11}^l}{\partial x^2} - \frac{\partial \Gamma_{12}^l}{\partial x^1} + \Gamma_{11}^s\Gamma_{2s}^l - \Gamma_{12}^s\Gamma_{1s}^l\right),$$

*so the Gaussian curvature is completely determined by the metric.*

Theorem 3.6 implies that isometric surfaces have the same Gaussian curvature. In particular, the following statement holds.

**Corollary 3.2.** *Suppose a domain $U$ on a surface can be mapped isometrically to a domain of the plane. Then the Gaussian curvature at each point of $U$ is equal to zero.*

**Proof.** Take the Euclidean coordinates $x$ and $y$ to be the coordinates on the surface. Since the mapping is isometric, the metric of the surface has the same form $dx^2 + dy^2$ as the metric of the plane. Hence the Gaussian curvature of the surface, being equal to that of the plane, is equal to zero. □

## Exercises to Chapter 3

1. Consider a surface of revolution given by

$$r(u, \varphi) = (u, f(u)\cos\varphi, f(u)\sin\varphi).$$

Find the first and the second fundamental forms and show that the meridians $\{\varphi = \text{const}\}$ and the parallels $\{u = \text{const}\}$ are mutually orthogonal. Consider, in particular, the cases of a torus and an ellipsoid of revolution.

2. Derive the parametric equation of the *cylindrical* surface formed by all the straight lines that are parallel to a vector $\xi$ and pass through a directrix curve $r = r(t)$.

3. Derive the parametric equation of the *conic* surface formed by all the straight lines which pass through a directrix curve $r = r(t)$ and a given point $P$ lying outside this curve.

4. A surface consisting of tangents to a curve is said to be *developable.* Derive the parametric equation of such a surface corresponding to a curve $r = r(t)$.

5. Derive the first and the second fundamental forms for the surface of revolution of the curve $y = a\cosh\frac{x}{a}$ about the axis $Ox$ (a catenoid) and find its principal curvatures.

6. A *ruled surface* is given parametrically by an equation of the form $r = r(u, v) = \rho(u) + va(u)$. Derive the equations of the ruled surface whose elements are straight lines parallel to the plane $y = z$ and intersecting the parabolas $y^2 = 2px$, $z = 0$ and $z^2 = -2px$, $y = 0$.

7. Show that the sum of angles of a triangle formed by arcs of great circles on a standard two-dimensional sphere in $\mathbb{R}^3$ is greater than $\pi$, and express the sum of angles in terms of the area of the triangle.

8. Derive the first and the second fundamental forms of the helicoid

$$r(u, v) = (u\cos v, u\sin v, av).$$

Prove that this surface is minimal (i.e., its mean curvature is identically zero) and find its principal curvatures.

9. Show that the surface of revolution

$$r(u, \varphi) = (f(u), u\cos\varphi, u\sin\varphi),$$

where

$$f(u) = \pm\Bigl(a\log\frac{a + \sqrt{a^2 - u^2}}{a} - \sqrt{a^2 - u^2}\Bigr), \quad a > 0,$$

has constant Gaussian curvature $K \equiv -1$ (this surface is referred to as the *Beltrami surface*).

10. Describe the surfaces all of whose normals intersect at the same point.

11. Find the Gaussian and the mean curvatures of the graph of a function

$$z = f(x) + g(y).$$

12. Show that a surface whose Gaussian and mean curvatures vanish everywhere is a plane.

13. Show that the mean curvature of the graph of a function $z = f(x, y)$ equals

$$H = \operatorname{div}\Bigl(\frac{\operatorname{grad} f}{\sqrt{1 + |\operatorname{grad} f|^2}}\Bigr).$$

14. Show that if the first fundamental form of a surface is equal to

$$dl^2 = A^2\,du^2 + B^2\,dv^2, \quad A = A(u,v), \quad B = B(u,v),$$

then the Gaussian curvature has the form

$$K = -\frac{1}{AB}\Big[\Big(\frac{A_v}{B}\Big)_v + \Big(\frac{B_u}{A}\Big)_u\Big].$$

15. Prove that two surfaces with the same constant Gaussian curvature are locally isometric.

16. Prove that cylindrical and conic surfaces are locally isometric to the plane.

17. Show that the metric on a surface of revolution has the form $dl^2 = du^2 + G(u)\,dv^2$ in appropriate coordinates. Find these coordinates for a sphere, torus, catenoid, and Beltrami surface.

18. For a surface $S$ formed by the tangent lines to a curve $\gamma$, express the Gaussian and mean curvatures of this surface in terms of the curvature and torsion of the curve $\gamma$.

19. Find the asymptotic lines on the following surfaces:

a) $z = \left(\frac{x}{y} + \frac{y}{x}\right)$;

b) $z = xy$.

20. Prove that there exist orthogonal asymptotic directions at a point of a surface if and only if the mean curvature at this point is equal to zero.

21. Prove that a two-dimensional pseudo-Riemannian metric of type $(1,1)$ with analytic coefficients can be reduced by a change of coordinates to the form $dl^2 = \lambda(x,y)(dx^2 - dy^2)$.

*Chapter 4*

# Complex Analysis in the Theory of Surfaces

## 4.1. Complex spaces and analytic functions

**4.1.1. Complex vector spaces.** A vector space of dimension $n$ over the field of complex numbers $\mathbb{C}$ consists of all vectors of the form

$$\xi = z^k e_k, \qquad z^k = x^k + iy^k,$$

where $z^k$ are complex coordinates, $e_1, \dots, e_n$ is a basis of the space, and $i$ is the imaginary unit ($i^2 = -1$). As in the real case, these vectors can be added,

$$\xi_1 = z_1^k e_k,\ \xi_2 = z_2^k e_k \ \rightarrow\ \xi_1 + \xi_2 = (z_1^k + z_2^k) e_k,$$

and multiplied by numbers, this time including complex numbers,

$$\lambda\xi = (\lambda z^k) e_k, \quad \lambda \in \mathbb{C}.$$

This space can also be regarded as a $2n$-dimensional vector space over the field of real numbers, with basis $e_1, \dots, e_n, ie_1, \dots, ie_n$:

$$\xi = z^k e_k = x^k e_k + y^k (ie_k).$$

The passage from a complex to a real vector space is called *realization.*

Linear transformations of a complex vector space are mappings specified by matrices $A = (a_j^k)$, $a_j^k \in \mathbb{C}$ and $\det A \neq 0$, which act by the formula

$$\xi \rightarrow A\xi : \quad \xi = \xi^k e_k, \quad A\xi = a_j^k \xi^j e_k.$$

This definition mimics the one for real vector spaces, and linear mappings of one space into another are defined similarly.

It follows from the definition that the linear transformations of the $n$-dimensional space form the group $\mathrm{GL}(n, \mathbb{C})$. This is the group of complex $(n \times n)$ matrices with nonzero determinant. The matrices in $\mathrm{GL}(n, \mathbb{C})$ with determinant equal to one form the group $\mathrm{SL}(n, \mathbb{C})$.

The vectors $\xi = z^k e_k$ are radius-vectors of points in the Cartesian space $\mathbb{C}^n$. The realization turns linear transformations of $\mathbb{C}^n$ into linear transformations of $\mathbb{R}^{2n}$. Thus we obtain the group homomorphism

$$r\colon \mathrm{GL}(n, \mathbb{C}) \to \mathrm{GL}(2n, \mathbb{R}),$$

which is called the realization chart. In order to distinguish between the real and complex cases, we write $\mathrm{GL}(k, \mathbb{R})$ for the group $\mathrm{GL}(k)$ of linear transformations of the space $\mathbb{R}^k$ (this group was introduced in Section 1.3.2).

Let us write down a matrix $A \in \mathrm{GL}(n, \mathbb{C})$ as the sum of its real and imaginary parts,

$$A = A_R + iA_I,$$

where $A_R, A_I$ are real $(n \times n)$ matrices. Upon realization, the matrix $A$ becomes

$$r(A) = \begin{pmatrix} A_R & -A_I \\ A_I & A_R \end{pmatrix}.$$

For example, the group $\mathrm{GL}(1, \mathbb{C})$ is formed by nonzero complex numbers with multiplication. Let $z = x + iy$ be the coordinate of an element of $\mathbb{C}$ and let $\lambda = a + ib \in \mathrm{GL}(1, \mathbb{C})$. Then

$$r(\lambda) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

The image of the mapping $r\colon \mathrm{GL}(n, \mathbb{C}) \to \mathrm{GL}(2n, \mathbb{R})$ is described by the following condition. Upon realization, multiplication by $i$ takes the form

$$i(e_k) = ie_k, \qquad i(ie_k) = -e_k.$$

Hence it is specified by the matrix

$$r(i) = \begin{pmatrix} 0 & -1_n \\ 1_n & 0 \end{pmatrix},$$

where $1_n$ is the identity matrix in $\mathrm{GL}(n, \mathbb{R})$. Since multiplication by $i$ commutes with any transformation $A \in \mathrm{GL}(n, \mathbb{C})$,

$$iA\xi = A(i\xi),$$

we obtain

$$r(i)r(A) = r(A)r(i).$$

Therefore, the matrices $r(A)$ commute with $r(i)$. It can be verified by a simple calculation that the matrices which commute with $r(i)$ have the block structure $\left(\begin{smallmatrix} B & -C \\ C & B \end{smallmatrix}\right)$, i.e., they exactly coincide with the image of the transformation $r$.

**4.1.2. The Hermitian scalar product.** Choose a scalar product such that the coordinates $x^1, \dots, x^n, y^1, \dots, y^n$ are Euclidean. The length of a curve specified in complex coordinates as $z^k(t) = x^k(t) + iy^k(t)$, $k = 1, \dots, n$, is equal to

$$\int \sqrt{\sum_{j=1}^{n}((\dot{x}^j)^2 + (\dot{y}^j)^2)}\, dt = \int \sqrt{\sum_{k=1}^{n} \dot{z}^k \dot{\bar{z}}^k}\, dt$$

(with the bar indicating the complex conjugate). Thus, trying to define the length of a vector in complex coordinates, we arrive at the complex-valued scalar product in $\mathbb{C}^n$:

$$\langle \xi_1, \xi_2 \rangle = \sum_{k=1}^{n} \xi_1^k \bar{\xi}_2^k. \tag{4.1}$$

It has the following properties:

1) $\langle \lambda\xi, \eta \rangle = \lambda \langle \xi, \eta \rangle$, $\langle \xi, \lambda\eta \rangle = \bar{\lambda} \langle \xi, \eta \rangle$;
2) $\langle \xi, \eta \rangle = \overline{\langle \eta, \xi \rangle}$;
3) $\langle \xi_1 + \xi_2, \eta \rangle = \langle \xi_1, \eta \rangle + \langle \xi_2, \eta \rangle$;
4) $\langle \xi, \xi \rangle > 0$ for $\xi \neq 0$.

Any scalar product with properties 1)–3) is said to be *Hermitian*. If, in addition, property 4) holds, then the scalar product is called *positive definite*.

One can prove, in a quite similar manner to the Euclidean case, that for any positive definite Hermitian scalar product there exists a basis in which it has the form (4.1).

We write down the Hermitian scalar product (4.1) in terms of real coordinates $x^j$, $y^k$, where $\xi_\alpha^j = x^j + iy^j$, $\alpha = 1, 2$, $j = 1, \dots, n$:

$$\langle \xi_1, \xi_2 \rangle = \sum_{k=1}^{n} (x_1^k + iy_1^k)(x_2^k - iy_2^k) = \sum_{k=1}^{n} (x_1^k x_2^k + y_1^k y_2^k) + i \sum_{k=1}^{n} (y_1^k x_2^k - x_1^k y_2^k).$$

Thus we have proved the following lemma.

**Lemma 4.1.** *A positive definite Hermitian scalar product in $\mathbb{C}^n$ has the form*

$$\langle \xi_1, \xi_2 \rangle = \langle \xi_1, \xi_2 \rangle_{\mathbb{R}} + i\omega(\xi_1, \xi_2),$$

*where $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ is a real symmetric (Euclidean) scalar product in $\mathbb{R}^{2n}$, and $\omega(\cdot, \cdot)$ is a symplectic scalar product in $\mathbb{R}^{2n}$. In particular, the Hermitian and real squared lengths of vectors coincide:*

$$\langle \xi, \xi \rangle = \langle \xi, \xi \rangle_{\mathbb{R}}.$$

REMARK. A similar result holds for all Hermitian scalar products, not necessarily positive definite.

Observe that

$$\langle i\xi_1, \xi_2\rangle = i\langle \xi_1, \xi_2\rangle = i\langle \xi_1, \xi_2\rangle_{\mathbb{R}} - \omega(\xi_1, \xi_2),$$

and, on the other hand,

$$\langle i\xi_1, \xi_2\rangle = \langle i\xi_1, \xi_2\rangle_{\mathbb{R}} + i\omega(i\xi_1, \xi_2).$$

Comparing these two expressions, we obtain

$$\langle \xi_1, \xi_2\rangle_{\mathbb{R}} = \omega(i\xi_1, \xi_2). \tag{4.2}$$

Therefore, the Hermitian scalar product is completely determined by its imaginary part, by the symplectic form $\omega$, or by its real part, which is a symmetric scalar product.

**4.1.3. Unitary and linear-fractional transformations.** A linear transformation $A \in \mathrm{GL}(n, \mathbb{C})$ is said to be *unitary* if it preserves the positive definite scalar product (4.1):

$$\langle A\xi_1, A\xi_2\rangle = \langle \xi_1, \xi_2\rangle$$

for all vectors $\xi_1$, $\xi_2$.

Choose the coordinates relative to which the Hermitian scalar product has the form (4.1). Then the condition for $A$ to be unitary is written as

$$\sum_{j=1}^{n} a_j^k \bar{a}_j^l = \delta^{kl},$$

or, in the matrix notation,

$$A^\top \bar{A} = 1. \tag{4.3}$$

The matrices satisfying relation (4.3) constitute the *unitary group* denoted by $\mathrm{U}(n)$. Equality (4.3) implies that

$$\det(A^\top \bar{A}) = \det A\,\overline{\det A} = |\det A|^2 = 1.$$

The subgroup consisting of unitary matrices with determinant equal to 1 is denoted by $\mathrm{SU}(n)$.

The group $\mathrm{U}(1) \subset \mathrm{GL}(1, \mathbb{C})$ consists of complex numbers $\lambda$ such that $|\lambda| = 1$. This means that $\lambda = e^{i\varphi} = \cos\varphi + i\sin\varphi$ for an appropriate angle $\varphi$, and the transformation

$$r(\lambda) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

is a rotation. This shows that the group $\mathrm{U}(1)$ is isomorphic to $\mathrm{SO}(2)$: realization turns $e^{i\varphi} \in \mathrm{U}(1)$ into rotation through the angle $\varphi$.

The Hermitian scalar product is completely determined by its real part via formula (4.2). Therefore, a transformation $A \in \mathrm{GL}(n,\mathbb{C})$ is unitary if and only if the transformation $r(A)$ is orthogonal,

$$r(\mathrm{U}(n)) = \mathrm{O}(2n) \cap r(\mathrm{GL}(n,\mathbb{C})).$$

The spaces $\mathbb{C}^n_{p,q}$ are defined analogously to the pseudo-Euclidean ones: in these spaces the squared length of a vector $\xi$ with coordinates $(\xi^1,\dots,\xi^n)$ is

$$\langle \xi,\xi\rangle = |\xi^1|^2 + \dots + |\xi^p|^2 - |\xi^{p+1}|^2 - \dots - |\xi^n|^2,$$

where $n = p+q$. The linear transformations preserving this form constitute the group $\mathrm{U}(p,q)$. The subgroup of $\mathrm{U}(p,q)$ consisting of all matrices with unit determinant is denoted by $\mathrm{SU}(p,q)$.

Now we will give more examples of transformation groups. A matrix $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$ belongs to $\mathrm{U}(2)$ if and only if

$$|a|^2 + |c|^2 = |b|^2 + |d|^2 = 1, \quad a\bar{b} + c\bar{d} = 0.$$

The subgroup $\mathrm{SU}(2)$ is specified by the additional condition $ad - bc = 1$; hence it consists of matrices of the form

$$\begin{pmatrix} a & b \\ -\bar{b} & \bar{a}\end{pmatrix}, \quad |a|^2 + |b|^2 = 1.$$

In its turn, the group $\mathrm{SU}(1,1)$ consists of the matrices

$$\begin{pmatrix} a & b \\ \bar{b} & \bar{a}\end{pmatrix}, \quad |a|^2 - |b|^2 = 1.$$

Define the mapping $\mathrm{SU}(1,1) \to \mathrm{SL}(2,\mathbb{R})$ by the formula

$$\begin{pmatrix} a & b \\ \bar{b} & \bar{a}\end{pmatrix} \to \begin{pmatrix} a_R + b_I & a_I + b_R \\ -a_I + b_R & a_R + b_I\end{pmatrix},$$

where $a = a_R + ia_I$, $b = b_R + ib_I$. It determines a group isomorphism $\mathrm{SU}(1,1) \simeq \mathrm{SL}(2,\mathbb{R})$.

Now, there is a homomorphism of the group $\mathrm{SL}(2,\mathbb{C})$ onto the group $L$ of linear-fractional transformations of the extended complex plane $\overline{\mathbb{C}} = \mathbb{C} \cup \infty$, i.e., the complex plane completed by the point at infinity. Namely, with each matrix $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$, $ad - bc = 1$, we associate the *linear-fractional transformation*

$$z \to z' = \frac{az+b}{cz+d}.$$

The composition of two linear-fractional transformations has the form

$$z \to z'' = \frac{a'z' + b'}{c'z' + d'} = \frac{(aa' + b'c)z + (a'b + b'd)}{(c'a + d'c)z + (c'b + d'd)}.$$

Hence the mapping

$$\mathrm{SL}(2,\mathbb{C}) \to L$$

constructed above is a homomorphism. It can be easily seen that its image is the entire group $L$, while the kernel consists of two matrices 1 and $-1$. Thus we have proved the following theorem.

**Theorem 4.1.** *The group $L$ of linear-fractional transformations of the extended complex plane $\overline{\mathbb{C}} = \mathbb{C} \cup \infty$ is isomorphic to the group* $\mathrm{SL}(2, \mathbb{C})/\pm 1$,

$$L \simeq \mathrm{SL}(2, \mathbb{C})/\pm 1.$$

**4.1.4. Holomorphic functions and the Cauchy–Riemann equations.** The passage to complex coordinates gives rise to partial derivatives

$$\frac{\partial}{\partial z^k} = \frac{1}{2}\Big(\frac{\partial}{\partial x^k} - i\,\frac{\partial}{\partial y^k}\Big),$$
$$\frac{\partial}{\partial \bar{z}^k} = \frac{1}{2}\Big(\frac{\partial}{\partial x^k} + i\,\frac{\partial}{\partial y^k}\Big),$$

which satisfy the natural conditions

$$\frac{\partial}{\partial z^j}\, z^k = \frac{\partial}{\partial \bar{z}^j}\, \bar{z}^k = \delta^k_j, \qquad \frac{\partial}{\partial z^j}\, \bar{z}^k = \frac{\partial}{\partial \bar{z}^j}\, z^k = 0.$$

Obviously,

$$\frac{\partial}{\partial x^k} = \frac{\partial}{\partial z^k} + \frac{\partial}{\partial \bar{z}^k}, \qquad \frac{\partial}{\partial y^k} = i\Big(\frac{\partial}{\partial z^k} - \frac{\partial}{\partial \bar{z}^k}\Big).$$

These formulas imply that the differential of a complex-valued function $f(z, \bar{z}) = f(x^1, \dots, x^n, y^1, \dots, y^n)$ has the form

$$df = \frac{\partial f}{\partial z^1}\, dz^1 + \dots + \frac{\partial f}{\partial z^n}\, dz^n + \frac{\partial f}{\partial \bar{z}^1}\, d\bar{z}^1 + \dots + \frac{\partial f}{\partial \bar{z}^n}\, d\bar{z}^n,$$

where

$$dz^k = dx^k + i\, dy^k, \qquad d\bar{z}^k = dx^k - i\, dy^k.$$

The operators $\frac{\partial}{\partial z^k}$ and $\frac{\partial}{\partial \bar{z}^k}$ are linear combinations with constant coefficients of the operators $\frac{\partial}{\partial x^k}$ and $\frac{\partial}{\partial y^k}$; hence they satisfy the Leibniz formula

$$\frac{\partial}{\partial z^k}\,(fg) = \frac{\partial f}{\partial z^k}\, g + f\, \frac{\partial g}{\partial z^k},$$
$$\frac{\partial}{\partial \bar{z}^k}\,(fg) = \frac{\partial f}{\partial \bar{z}^k}\, g + f\, \frac{\partial g}{\partial \bar{z}^k}.$$

A function $f(x^1, \dots, x^n, y^1, \dots, y^n)$ is said to be *complex-analytic* or *holomorphic* if

$$\frac{\partial f}{\partial \bar{z}^k} \equiv 0, \quad k = 1, \dots, n.$$

The simplest example of a complex-analytic function is a polynomial $P(z)$.

A polynomial $P(z, \bar{z})$ of two variables can be written as a polynomial of $z$ alone if and only if

$$\frac{\partial P}{\partial \bar{z}} = 0.$$

Indeed, let $P(z, \bar{z}) = a_0 z^m + \cdots + a_{m-1} z + a_m$, where $a_0, \ldots, a_m$ are constants. Since

$$\frac{\partial}{\partial \bar{z}}[z^m] = 0, \qquad \frac{\partial}{\partial \bar{z}}[\bar{z}^m] = m\bar{z}^{m-1},$$

we obtain that $\partial P/\partial \bar{z} \equiv 0$. Conversely, any nonzero polynomial $P(z, \bar{z})$ can be written as

$$P(z, \bar{z}) = a_0 \bar{z}^m + \cdots + a_{m-1}\bar{z} + a_m,$$

where $a_0, \ldots, a_m$ are polynomials in $z$ and the leading coefficient $a_0$ is not identically zero. If $m \neq 0$, then

$$\frac{\partial P}{\partial \bar{z}} = a_0 \cdot m\bar{z}^{m-1} + \cdots$$

and $a_0 \neq 0$ implies $\frac{\partial P}{\partial \bar{z}} \neq 0$.

This fact holds also for convergent power series.

For $n = 1$ the analyticity condition becomes

$$\frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) = 0.$$

Decomposing $f$ into the real and imaginary parts, $f(x, y) = u(x, y) + iv(x, y)$, we can rewrite these conditions in the form of the *Cauchy–Riemann equations*:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

These equations imply that

$$\Delta u = \Delta v = 0,$$

where

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = 4\frac{\partial^2}{\partial z \partial \bar{z}}$$

is the *Laplace operator*. The functions $\varphi$ satisfying the Laplace equation $\Delta\varphi = 0$ are called *harmonic*.

As we have shown, the real and imaginary parts of a complex-analytic function are harmonic functions.

**4.1.5. Complex-analytic coordinate changes.** A set $U$ in $\mathbb{C}^n$ is called a domain in $\mathbb{C}^n$ if it is a domain in the realized space $\mathbb{R}^{2n}$. Suppose that there are two complex coordinate systems in the domain $U$:

$$z^1 = x^1 + iy^1, \dots, z^n = x^n + iy^n,$$
$$w^1 = u^1 + iv^1, \dots, w^n = u^n + iv^n.$$

This means that there are smooth functions

$$w^k = w^k(x^1, \dots, x^n, y^1, \dots, y^n), \quad k = 1, \dots, n, \tag{4.4}$$

such that the mapping $(x, y) \to (u, v)$ is invertible and the inverse mapping is also smooth.

The change of coordinates (4.4) is called complex-analytic (holomorphic) if

$$\frac{\partial w^k}{\partial \bar{z}^l} \equiv 0, \quad k, l = 1, \dots, n.$$

The complex Jacobi matrix for such a change of coordinates is defined as

$$a_l^k = \frac{\partial w^k}{\partial z^l},$$

and its determinant is called the complex Jacobian:

$$J_{\mathbb{C}} = \det(a_l^k).$$

The change (4.4) determines at the same time a change of coordinates in the realized space:

$$u^k = u^k(x^1, \dots, x^n, y^1, \dots, y^n),$$
$$v^k = v^k(x^1, \dots, x^n, y^1, \dots, y^n), \quad k = 1, \dots, n.$$

Denote its Jacobian by $J_{\mathbb{R}}$:

$$J_{\mathbb{R}} = \det \frac{\partial(u, v)}{\partial(x, y)}.$$

**Lemma 4.2.** *For a complex-analytic change of coordinates,*

$$J_{\mathbb{R}} = |J_{\mathbb{C}}|^2.$$

**Proof.** Since

$$\frac{\partial w^j}{\partial z^k} = a_k^j, \qquad \frac{\partial \bar{w}^j}{\partial \bar{z}^k} = \bar{a}_k^j, \qquad \frac{\partial w^j}{\partial \bar{z}^k} = \frac{\partial \bar{w}^j}{\partial z^k} = 0,$$

the Jacobi matrix of transition from the coordinates $z^1, \dots, z^n$, $\bar{z}^1, \dots, \bar{z}^n$ to the coordinates $w^1, \dots, w^n$, $\bar{w}^1, \dots, \bar{w}^n$ equals

$$\begin{pmatrix} A & 0 \\ 0 & \bar{A} \end{pmatrix}.$$

Its determinant is

$$\det A \cdot \det \bar{A} = |\det A|^2 = |J_{\mathbb{C}}|^2.$$

The transition from the coordinates $x^1, \dots, x^n, y^1, \dots, y^n$ to the coordinates $z^1, \dots, z^n, \bar{z}^1, \dots, \bar{z}^n$ or from $u^1, \dots, u^n, v^1, \dots, v^n$ to $w^1, \dots, w^n, \bar{w}^1, \dots, \bar{w}^n$ is a linear mapping specified by the matrix

$$B = \begin{pmatrix} 1_n & i \cdot 1_n \\ 1_n & -i \cdot 1_n \end{pmatrix}, \quad \det B = (-2i)^n,$$

where $1_n$ is the $(n \times n)$ identity matrix.

Now we consider the composition of the coordinate changes $(x, y) \to (z, \bar{z}) \to (w, \bar{w}) \to (u, v)$ to obtain

$$J_{\mathbb{R}} = \det\left( B^{-1} \begin{pmatrix} A & 0 \\ 0 & \bar{A} \end{pmatrix} B \right) = |J_{\mathbb{C}}|^2.$$

This proves the lemma. □

The proof of Lemma 4.2 and the inverse function theorem entail the following assertion.

**Lemma 4.3.** *If the Jacobian of a complex-analytic change*

$$w^k = w^k(z^1, \dots, z^n), \quad k = 1, \dots, n,$$

*does not vanish, then this change is locally (in a neighborhood of each point) invertible:*

$$z^k = z^k(w^1, \dots, w^n), \quad k = 1, \dots, n,$$

*with complex-analytic functions $z^k(w)$.*

**Proof.** The existence of the functions $z^k(w, \bar{w})$ follows from the inverse function theorem. The Jacobi matrix of the change $(w, \bar{w}) \to (z, \bar{z})$ has the form

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & \bar{A}^{-1} \end{pmatrix}.$$

Therefore, $\partial z^j / \partial \bar{w}^k \equiv 0$ for $j, k = 1, \dots, n$, which proves the lemma. □

Let $U$ be a domain on a surface with coordinates $x$ and $y$. Then $z = x+iy$ determines a complex coordinate (or *complex parameter*) on the surface. This parameter is said to be *conformal* if the first fundamental form writes as

$$g(x, y)(dx^2 + dy^2) = g(z, \bar{z})\, dz\, d\bar{z}. \tag{4.5}$$

In this case the coordinates $x$ and $y$ are also called *conformal* or *isothermal.*

**Lemma 4.4.** *The property* (4.5) *of a metric to be conformal is invariant with respect to the class of complex-analytic changes of coordinates and their compositions with complex conjugation.*

**Proof.** Let $z = z(w)$ and $\partial z/\partial \bar{w} = 0$. Then

$$dz = \left(\frac{dz}{dw}\right) dw, \qquad d\bar{z} = \overline{\left(\frac{dz}{dw}\right)}\, d\bar{w}$$

and

$$g(z, \bar{z})\, dz\, d\bar{z} = g(w, \bar{w}) \left|\frac{dz}{dw}\right|^2 dw\, d\bar{w}.$$

The proof for $\partial z/\partial w = 0$ is similar.

Therefore the transformations stated in the lemma preserve the conformal metric.

Now let $z = z(w, \bar{w})$, $\partial z/\partial w \neq 0$, and $\partial z/\partial \bar{w} \neq 0$ in some domain $U$. Then

$$\begin{aligned} g(z, \bar{z})\, dz\, d\bar{z} &= g(w, \bar{w}) \Big(\frac{\partial z}{\partial w}\, dw + \frac{\partial z}{\partial \bar{w}}\, d\bar{w}\Big)\Big(\frac{\partial \bar{z}}{\partial w}\, dw + \frac{\partial \bar{z}}{\partial \bar{w}}\, d\bar{w}\Big) \\ &= g(w, \bar{w}) \Big(\frac{\partial z}{\partial w}\frac{\partial \bar{z}}{\partial w}\, dw^2 + \Big(\frac{\partial z}{\partial w}\frac{\partial \bar{z}}{\partial \bar{w}} + \frac{\partial z}{\partial \bar{w}}\frac{\partial \bar{z}}{\partial w}\Big) dw\, d\bar{w} + \frac{\partial z}{\partial \bar{w}}\frac{\partial \bar{z}}{\partial \bar{w}}\, d\bar{w}^2\Big) \\ &= \tilde{E}\, du^2 + 2\tilde{F}\, du\, dv + \tilde{G}\, dv^2, \end{aligned}$$

where $w = u + iv$ and the coefficient of $dw^2$ does not vanish in the domain $U$. The proof is completed. □

## 4.2. Geometry of the sphere

**4.2.1. The metric of the sphere.** The sphere $S^2 \subset \mathbb{R}^3$ of radius $R$ with center at the origin is specified by the equation

$$x^2 + y^2 + z^2 = R^2. \tag{4.6}$$

In spherical coordinates $r$, $\theta$, $\varphi$ this equation takes a particularly simple form

$$r = R.$$

Hence $\theta$ and $\varphi$ parametrize the sphere except for two points, the "North and South Poles". At these points $\theta = 0$ and $\theta = \pi$, respectively.

In spherical coordinates the Euclidean metric in $\mathbb{R}^3$ has the form

$$dx^2 + dy^2 + dz^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2).$$

On a level surface $r = R$ the differential $dr$ vanishes, and the metric of the sphere of radius $R$ becomes

$$dl^2 = R^2(d\theta^2 + \sin^2\theta\, d\varphi^2),$$

where $0 \le \varphi \le 2\pi$, $0 < \theta < \pi$.

The *distance* between two points $P$ and $Q$ on a surface is defined as the infimum

$$\rho(P, Q) = \inf l(r)$$

of the lengths of the curves $r$ on the surface joining the points $P$ and $Q$.

Let $P_+$ be the "North Pole" (i.e., $\theta = 0$) and $Q = (\theta_0, \varphi_0)$. Consider a curve $r\colon [a,b] \to S^2$ joining these points: $r(a) = P_+$ and $r(b) = Q$. Then we have

$$l(r) = \int_a^b R\sqrt{\dot\theta^2 + \sin^2\theta\,\dot\varphi^2}\,dt \geq \int_a^b R\sqrt{\dot\theta^2}\,dt \geq \int_0^{\theta_0} R\,d\theta = R\theta_0.$$

A great circle on a sphere $S^2$ is obtained by intersecting the sphere with a plane passing through the origin. It is easily seen that the curve of minimal length among those joining $P_+$ and $Q$ is the smaller part of the great circle. Indeed, since $l(r) = R\theta_0$ only for $\dot\varphi \equiv 0$, the length of any other curve is greater than $R\theta_0$.

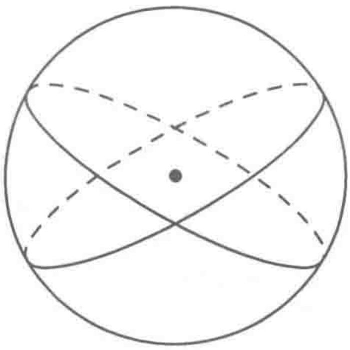

**Figure 4.1.** Great circles: geodesics on a sphere.

Thus we have proved the following lemma.

**Lemma 4.5.** *The distance $\rho(P_+, Q)$ from the "North Pole" $P_+ = (0, 0, R)$ to a point $Q = (\theta_0, \varphi_0)$ is equal to $R\theta_0$ and is attained exactly on the smaller part of the corresponding great circle.*

Denote by $B_\varepsilon$ the circle of radius $\varepsilon$ with center $P_+$ on the sphere, i.e., the set of points $Q$ such that $\rho(P_+, Q) \leq \varepsilon$. In spherical coordinates it is specified by the inequality

$$\theta \leq \frac{\varepsilon}{R},$$

and its boundary (the circumference) is given by the equation

$$\theta = \frac{\varepsilon}{R}.$$

The length of this circumference is equal to

$$l_\varepsilon = \int_0^{2\pi} R\sin\frac{\varepsilon}{R}\,d\varphi = 2\pi R\sin\frac{\varepsilon}{R},$$

and the area of the circle $B_\varepsilon$ is equal to

$$\sigma_\varepsilon = \int_0^{2\pi} d\varphi \int_0^{\varepsilon/R} R^2\sin\alpha\,d\alpha = 2\pi R^2\Big(1-\cos\frac{\varepsilon}{R}\Big).$$

The length $l_\varepsilon$ attains its maximum at the equator $\theta = \pi/2$, where it is equal to $2\pi R$, i.e., to the length of the great circle. For $\varepsilon = \pi R$ the circle $B_\varepsilon$ covers the entire sphere, and we see that the area of the sphere is $4\pi R^2$.

As $\varepsilon \to 0$, we have the following expansions:

$$l_\varepsilon = 2\pi\varepsilon - \frac{\pi}{3R^2}\,\varepsilon^3 + O(\varepsilon^5), \qquad \sigma_\varepsilon = \pi\varepsilon^2 - \frac{\pi}{12R^2}\,\varepsilon^4 + O(\varepsilon^6). \tag{4.7}$$

Observe that the normal sections of the sphere are great circles. They have a constant curvature equal to $R^{-1}$. Therefore, the Gaussian and mean curvatures of the sphere are

$$K = \frac{1}{R^2}, \qquad H = \frac{1}{R}.$$

Then the expansions (4.7) can be restated as

$$l_\varepsilon = 2\pi\varepsilon - \frac{K}{3}\,\pi\varepsilon^3 + O(\varepsilon^5), \qquad \sigma_\varepsilon = \pi\varepsilon^2 - \frac{K}{12}\,\pi\varepsilon^4 + O(\varepsilon^6). \tag{4.8}$$

The leading terms of these expansions are the length of the circumference and the area of the circle of radius $\varepsilon$ on the Euclidean plane. The departures from the Euclidean case are estimated in terms of the curvature:

$$\Delta l_\varepsilon \simeq -\frac{K}{3}\,\pi\varepsilon^3, \qquad \Delta\sigma_\varepsilon \simeq -\frac{K}{12}\,\pi\varepsilon^4$$

as $\varepsilon \to 0$.

**4.2.2. The group of motions of a sphere.** Consider motions of a sphere $S^2$, i.e., transformations $\psi\colon S^2 \to S^2$ that preserve the lengths of curves:

$$l(r) = l(\psi(r))$$

for any curve $r$. It follows from the definition of the distance that the motions preserve the distances between points:

$$\rho(P, Q) = \rho(\psi(P), \psi(Q))$$

for any points $P, Q \in S^2$.

An orthogonal transformation $x \to Ax$, where $A \in \mathrm{O}(3)$, maps the sphere (4.6) into itself and preserves the Euclidean metric in $\mathbb{R}^3$. Hence any orthogonal transformation is a motion of the sphere.

For any pair of points $P$ and $Q$ on the sphere, there exists an orthogonal transformation $A$ such that $A(Q) = P$. In particular, any point can be carried into the "North Pole" $P_+ = (0, 0, R)$. Therefore, roughly speaking, the sphere has the same form in a neighborhood of any of its points. Moreover, orthogonal transformations map great circles again into great circles. Thus we have proved the following lemma.

**Lemma 4.6.** *The asymptotic formulas* (4.8) *for the length of the circumference and the area of a circle of radius $\varepsilon$ hold in a neighborhood of any point of the sphere.*

*The distance between any two points $P, Q \in S^2$ is equal to the length of the smaller part of the great circle on which these points lie. If $\rho(P, Q) < \pi R$, then there is a unique great circle which passes through the points $P$ and $Q$.*

Before completing the description of the group of motions of a sphere we will prove the following assertion, which is a corollary of Lemma 4.5.

**Lemma 4.7.** *If the point $P_+ = (0, 0, R)$ stays fixed under a motion of the sphere* (4.6), *then this motion is either a rotation about the $z$-axis or a composition of such rotation with reflection in the plane $xz$: $(x, y, z) \to (x, -y, z)$.*

**Proof.** Since a motion preserves distances between points, by Lemma 4.5 it maps the circles $\theta = \text{const}$ into themselves. It acts on them by the formula

$$\varphi \to \varepsilon(\theta)\varphi + \varphi_0(\theta),$$

where $\varepsilon(\theta) = \pm 1$. The motion maps the shorter part of the line $\varphi = \text{const}$ passing through $P_+$ again into the shorter part of such a line. Therefore, $\varepsilon(\theta)$ and $\varphi_0(\theta)$ do not depend on $\theta$, and the motion has the form

$$\varphi \to \pm\varphi + \varphi_0,$$

which proves the lemma. □

**Theorem 4.2.** *Any motion of the sphere* (4.6) *is induced by an orthogonal transformation of the space $\mathbb{R}^3$.*

**Proof.** Let $\psi \colon S^2 \to S^2$ be a motion and let $\psi(P_+) = P$. Consider an orthogonal transformation that maps $P$ into $P_+$. Then $\chi\psi$ is a motion keeping the point $P_+$ fixed. By Lemma 4.7 this motion is an orthogonal transformation. Therefore, $\psi = \chi^{-1}(\chi\psi)$ is an orthogonal transformation. □

**Corollary 4.1.** *The motions of the sphere $S^2$ form a group that is isomorphic to* O(3).

Let $P_+ = (0, 0, R)$ and $P_- = (0, 0, -R)$ be the "North and South Poles" of the sphere $S^2$. We construct the stereographic projections from these points onto the plane $xy$.

Let $Q \in S^2 \setminus P_+$ be a point of the sphere different from $P_+$. Draw a straight line through $Q$ and $P_+$ and denote by $\pi_+(Q)$ the intersection point of this line with the plane $xy$. The mapping

$$Q \to \pi_+(Q)$$

is called the *stereographic projection* from $P_+$. In polar coordinates $(r, \varphi)$ on the plane $xy$ it is written as

$$(\theta, \varphi) \to \left(R \cot \frac{\theta}{2}, \varphi\right).$$

This mapping is invertible, hence $(x, y)$ may be taken for coordinates on the sphere $S^2$ with deleted "North Pole" $P_+$. The metric $dl^2 = R^2(d\theta^2 + \sin^2\theta \, d\varphi^2)$ on the sphere is given in these coordinates by the formula

$$dl^2 = \frac{4R^4}{(R^2 + x^2 + y^2)^2} (dx^2 + dy^2).$$

Define a complex parameter

$$z_+ = \frac{x + iy}{R}$$

on $S^2 \setminus P_+$. The metric on the sphere in terms of this parameter becomes

$$dl^2 = \frac{4R^2}{(1 + z_+\bar{z}_+)^2} \, dz_+ \, d\bar{z}_+.$$

We see that the parameter $z_+$ is conformal.

The stereographic projection $\pi_-$ from the "South Pole" $P_-$ is defined in a similar manner. It is given by the formula

$$(\theta, \varphi) \to \left(R \tan \frac{\theta}{2}, \varphi\right).$$

On $S^2 \setminus P_-$ we can introduce the parameter

$$z_- = \frac{x - iy}{R},$$

where $(x, y) = \pi_-(Q)$. It is also conformal:

$$dl^2 = \frac{4R^2}{(1 + z_-\bar{z}_-)^2} \, dz_- \, d\bar{z}_-.$$

If $Q \in S^2 \setminus \{P_+, P_-\}$, i.e., $Q$ is a point on the sphere different from $P_+$ and $P_-$, then its coordinates $z_+$ and $z_-$ are related as

$$z_+ = \frac{1}{z_-}.$$

Hence the change of coordinates $z_+ \to z_-$ is complex-analytic. Thus we have proved the following theorem.

**Theorem 4.3.** *The sphere* (4.6) *is covered by the domains $U_+$ and $U_-$, where $U_\pm = S^2 \setminus P_\pm$, $S^2 = U_+ \cup U_-$. In these domains we have the conformal parameters $z_+$ and $z_-$ related by the formula*

$$z_+ z_- = 1$$

*in the intersection $U_+ \cap U_-$ of the domains. On $U_\pm$ the metric of the sphere has the form*

$$\frac{4R^2}{(1+|z_\pm|^2)^2}\, dz_\pm\, d\bar{z}_\pm.$$

It is natural to assign to the points $P_+$ and $P_-$ the values $z_+ = \infty$ and $z_- = \infty$ of the corresponding parameters. For this reason in complex geometry the sphere $S^2$ is identified with the extended complex plane $\overline{\mathbb{C}}$.

On $\overline{\mathbb{C}}$ the linear-fractional transformations

$$z \to w = \frac{az+b}{cz+d}, \quad ad - bc = 1,$$

are defined. Since

$$\frac{\partial w}{\partial z} = \frac{ad-bc}{(cz+d)^2} = \frac{1}{(cz+d)^2},$$

we have

$$dw\, d\bar{w} = \frac{dz\, d\bar{z}}{|cz+d|^4}.$$

Thus we obtain that linear-fractional transformations preserve the conformal form of the Euclidean metric.

Take $z$ equal to $z_+$ or $z_-$. Then any linear-fractional transformation maps the sphere $S^2$ onto itself. Let us find motions of the sphere among these transformations. They must satisfy the equality

$$\frac{dw\, d\bar{w}}{(1+w\bar{w})^2} = \frac{dz\, d\bar{z}}{(1+z\bar{z})^2}.$$

Rewrite the left-hand side as

$$\frac{1}{\left(1+\frac{|az+b|^2}{|cz+d|^2}\right)^2} \cdot \frac{dz\, d\bar{z}}{|cz+d|^4} = \frac{dz\, d\bar{z}}{(|cz+d|^2 + |az+b|^2)^2}$$

$$= \frac{dz\, d\bar{z}}{((|a|^2+|c|^2)z\bar{z} + (a\bar{b}+c\bar{d})z + (\bar{a}b+\bar{c}d)\bar{z} + (|b|^2+|d|^2))^2}.$$

Therefore, a linear-fractional transformation is a motion if

$$|a|^2+|c|^2 = 1, \qquad |b|^2+|d|^2 = 1, \qquad a\bar{b}+c\bar{d} = 0, \qquad ad-bc = 1.$$

These equalities hold if and only if the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ belongs to SU(2), i.e., $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} a & b \\ -\bar{b} & \bar{a} \end{smallmatrix}\right)$ and $|a|^2 + |b|^2 = 1$.

Thus we obtain a homomorphism

$$\mathrm{SU}(2) \to \mathrm{O}(3)$$

with kernel $\pm 1$. Indeed, since all motions of the sphere $S^2$ are induced by orthogonal transformations of $\mathbb{R}^3$ (Theorem 4.2), the identical transformation in O(3) keeps all the points of the sphere fixed; hence it may result only from the identical linear-fractional transformation $z \to z$, i.e., $\pm 1 \in \mathrm{SU}(2)$.

A motion of the sphere $S^2$ is called proper if it is induced by a proper orthogonal transformation $x \to Ax$. For $a = 1$, $b = 0$ we get the identical transformation, which is proper. Since $\det A$ is a continuous function of $a$ and $b$ and it can take only two values 1 and $-1$, linear-fractional transformations can produce only proper motions. Thus we conclude that the linear-fractional motions form a subgroup of the group of all proper motions of the sphere. The following theorem shows that this is actually the entire group.

**Theorem 4.4.** *The group* $\mathrm{SU}(2)/\pm 1$ *of linear-fractional motions of the sphere* (4.6) *is isomorphic to the group* SO(3) *of proper orthogonal transformations of the space* $\mathbb{R}^3$.

**Proof.** It is easily seen that different elements of $\mathrm{SU}(2)/\pm 1$ determine different linear-fractional motions. Hence it suffices to show that any motion in SO(3) is given by a linear-fractional transformation.

The projection $\pi_+$ maps the point $P_-$ into the point $z = 0$. All linear-fractional motions of the sphere $\overline{\mathbb{C}}$ are given by the formula

$$z \to \frac{az+b}{-\bar{b}z+\bar{a}}, \quad |a|^2 + |b|^2 = 1.$$

For $b = 0$ we obtain all rotations of the sphere leaving the point $P_-$ fixed.

Let $\psi \in \mathrm{SO}(3)$, $\psi(P_-) = P$, and $\pi_+(P) = z_0$. Linear-fractional motions can move the point $z = 0 \in \overline{\mathbb{C}}$ into any point of $\overline{\mathbb{C}}$. Take a linear-fractional motion $\chi$ such that $\chi(z_0) = 0$. The motion $\chi\psi$ is proper and leaves the point $P_-$ fixed. By Lemma 4.7 the mapping $\chi\psi$ is a rotation keeping the point $P_-$ fixed. Therefore, the motions $\chi\psi$ and hence $\psi$ are given by linear-fractional transformations. The proof is completed. □

## 4.3. Geometry of the pseudosphere

**4.3.1. Space-like surfaces in pseudo-Euclidean spaces.** Let $\mathbb{R}^{1,2}$ be the pseudo-Euclidean space with coordinates $(t, x, y)$ and metric

$$dl^2 = dt^2 - dx^2 - dy^2. \tag{4.9}$$

In the domain $t^2 - x^2 - y^2 > 0$ we introduce pseudospherical coordinates $(\rho, \chi, \varphi)$:

$$t = \rho \cosh \chi, \qquad x = \rho \sinh \chi \cos \varphi, \qquad y = \rho \sinh \chi \sin \varphi,$$

where $-\infty < \rho < \infty$, $0 < \chi < \infty$, and $0 \leq \varphi \leq 2\pi$. In these coordinates the metric becomes

$$dl^2 = d\rho^2 - \rho^2(d\chi^2 + \sinh^2 \chi \, d\varphi^2).$$

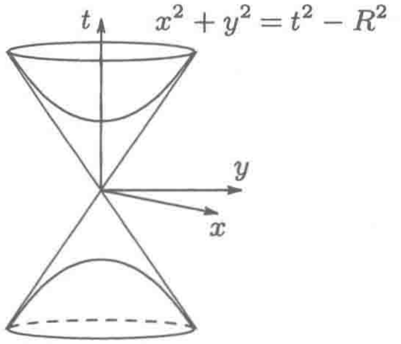


**Figure 4.2.** Pseudosphere.

A pseudosphere of radius $R$ is given by the equation

$$t^2 - x^2 - y^2 = R^2 \tag{4.10}$$

(see Figure 4.2). It is a two-sheeted hyperboloid, and its upper and lower parts are specified in pseudospherical coordinates by the equations

$$\rho = R \quad \text{and} \quad \rho = -R.$$

We will restrict our attention to the upper part, where $\rho = R$.

The pseudosphere is an example of a *space-like surface* in $\mathbb{R}^{1,2}$, which means that all tangent vectors to this surface are space-like. Hence, in order to make the induced metric positive definite, we must reverse its sign: by definition, the first fundamental form on such a surface is $-dl^2$.

There are two ways to define the Gaussian curvature of space-like surfaces in $\mathbb{R}^{1,2}$.

First, take a unit time-like normal vector $n$ to the surface and specify the second fundamental form by the equality

$$b_{ij} = \left\langle \frac{\partial^2 r}{\partial x^i \partial x^j}, n \right\rangle, \quad i, j = 1, 2,$$

where $r = r(x^1, x^2)$ is a parametrization of the surface. Now we let

$$K = -\frac{\det b_{ij}}{\det g_{ij}}.$$

For example, for a surface given as the graph of a function $t = f(x, y)$ we have

$$g_{ij}\, dx^i\, dx^j = -dl^2 = (1 - f_x^2)\, dx^2 - 2f_x f_y\, dx\, dy + (1 - f_y^2)\, dy^2,$$

where $x^1 = x$, $x^2 = y$. The normal vector is

$$n = \frac{(1, f_x, f_y)}{\sqrt{1 - f_x^2 - f_y^2}},$$

and by definition the Gaussian curvature equals

$$K = \frac{f_{xy}^2 - f_{xx} f_{yy}}{(1 - f_x^2 - f_y^2)^2}. \tag{4.11}$$

Secondly, the Gaussian curvature may be obtained from the metric using the formulas in Lemma 3.10 and Theorem 3.6.

By using analogs of the Codazzi equations for space-like surfaces in $\mathbb{R}^{1,2}$ one can show that both ways lead to the same result.

Recall that for surfaces in $\mathbb{R}^3$ we had a similar formula (3.2),

$$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2},$$

but with different signs.

The upper part of the pseudosphere is given by the equation

$$t = f(x, y) = \sqrt{R^2 + x^2 + y^2}.$$

Substituting this function into the formula for the Gaussian curvature in $\mathbb{R}^{1,2}$, we obtain

$$K \equiv -\frac{1}{R^2}.$$

Thus we have shown that the pseudosphere has a constant negative curvature.

**4.3.2. The metric and the group of motions of the pseudosphere.** On the pseudosphere $\rho = R$ we have $d\rho = 0$, so the induced metric is

$$-dl^2 = R^2(d\chi^2 + \sinh^2 \chi\, d\varphi^2).$$

This metric is called the *Lobachevsky metric* or the *hyperbolic metric.* For $R = 1$ the pseudosphere with this metric is called the *Lobachevsky plane* or *hyperbolic plane.*

The pseudoorthogonal transformations (elements of $\mathrm{O}(1, 2)$) map the pseudosphere into itself and, more specifically, they are its motions. Any point of the pseudosphere may be sent into any other point of the pseudosphere by such a transformation. Therefore, when evaluating the area and

the circumference of a circle of radius $r$ with center at a point $P$, we can take for $P$ the "North Pole" $P_+ = (R, 0, 0)$.

The proofs of the following assertions are similar to those of Lemmas 4.5, 4.7, and Theorem 4.2.

**Lemma 4.8.** *The distance between the point $P_+ = (R, 0, 0)$ and a point $Q$ of the upper part of the pseudosphere is $R\chi_0$, where $(\chi_0, \varphi_0)$ are the coordinates of the point $Q$. It is attained exactly on the part of the "straight line" $\varphi = \text{const}$ from $P_+$ to $Q$.*

**Lemma 4.9.** *If a motion of the pseudosphere* (4.10) *leaves the point $P_+ = (R, 0, 0)$ fixed, then it is either a rotation about the $y$-axis or the composition of such a rotation with the reflection in the plane $xy$: $(t, x, y) \to (-t, x, y)$.*

**Theorem 4.5.** *The class of motions of the pseudosphere* (4.10) *coincides with the class of pseudoorthogonal transformations of $\mathbb{R}^{1,2}$, i.e., of linear transformations preserving the metric* (4.9).

The circumference of a circle of radius $\varepsilon$ is equal to

$$l_\varepsilon = \int_0^{2\pi} R \sinh \frac{\varepsilon}{R}\, d\varphi = 2\pi R \sinh \frac{\varepsilon}{R},$$

and the area of this circle is

$$\sigma_\varepsilon = \int_0^{2\pi} d\varphi \int_0^{\varepsilon/R} R^2 \sinh \alpha \, d\alpha = 2\pi R^2 \Big(\cosh \frac{\varepsilon}{R} - 1\Big).$$

As $\varepsilon \to 0$, we have the following expansions:

$$l_\varepsilon = 2\pi\varepsilon + \frac{\pi}{3R^2}\,\varepsilon^3 + O(\varepsilon^5), \qquad \sigma_\varepsilon = \pi\varepsilon^2 + \frac{\pi}{12R^2}\,\varepsilon^4 + O(\varepsilon^6). \tag{4.12}$$

The expansions (4.12) can be rewritten as

$$l_\varepsilon = 2\pi\varepsilon - \frac{K}{3}\,\pi\varepsilon^3 + O(\varepsilon^5), \qquad \sigma_\varepsilon = \pi\varepsilon^2 - \frac{K}{12}\,\pi\varepsilon^4 + O(\varepsilon^6).$$

These formulas coincide with their analogs (4.8) for the sphere (4.6). As in that case, the leading terms of these expansions are the circumference and the area of a circle of radius $\varepsilon$ on the Euclidean plane, and deviations from the planar case are estimated in terms of the Gaussian curvature:

$$\Delta l_\varepsilon \simeq -\frac{K}{3}\,\pi\varepsilon^3, \qquad \Delta\sigma_\varepsilon \simeq -\frac{K}{12}\,\pi\varepsilon^4$$

as $\varepsilon \to 0$. Since $K = 0$ in the planar case, we conclude that these formulas hold for all curvatures.

**4.3.3. Models of hyperbolic geometry.** Let us construct the stereographic projection $\pi$ of the upper part of the pseudosphere from the "South Pole" $P_- = (-R, 0, 0)$ onto the plane $xy$. Let $Q$ be a point on the pseudosphere. We draw a line segment joining $Q$ with $P_-$. Denote by $\pi(Q)$ the intersection point of this segment with the plane $xy$ (see Figure 4.3). In polar coordinates on the plane this projection has the form

$$(\chi, \varphi) \to \Big(\frac{R\sinh\chi}{1+\cosh\chi}, \varphi\Big).$$

This mapping is invertible, its image is the interior of the circle $x^2+y^2 < R^2$, and hence $(x, y)$ can be taken for coordinates on the pseudosphere.

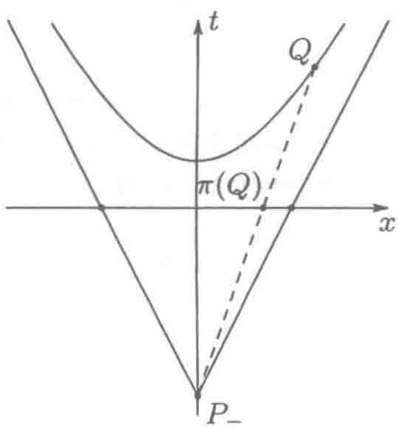


**Figure 4.3.** Stereographic projection on the pseudosphere.

In these coordinates the metric of the sphere, $R^2(d\chi^2 + \sinh^2\chi\, d\varphi^2)$, becomes

$$-dl^2 = \frac{4R^4}{(R^2 - x^2 - y^2)^2}(dx^2 + dy^2).$$

Defining a conformal parameter

$$z = \frac{x + iy}{R}$$

on the pseudosphere, we can write the metric of the pseudosphere as

$$-dl^2 = \frac{4R^2}{(1 - |z|^2)^2}\, dz\, d\bar{z},$$

where $|z| < 1$. This representation of the Lobachevsky metric on the pseudosphere for $R = 1$ is called *Poincaré's model* of the Lobachevsky plane.

Next, we perform the following one-to-one mapping of the circle $\{|z| < 1\}$ onto the half-plane $\{(x, y) \mid y > 0\}$:

$$z \to w = i\frac{1-z}{1+z}, \qquad w = x + iy.$$

The coordinates $(x, y)$ on the half-plane can again be taken for coordinates on the pseudosphere. In these coordinates the Lobachevsky metric becomes

$$R^2 \frac{dx^2 + dy^2}{y^2}.$$

The half-plane $y > 0$ with this metric provides another model of the Lobachevsky plane.

The curvature of the Lobachevsky plane is $K = -1$ everywhere (see Section 4.3.1).

A linear-fractional transformation

$$z \to w = \frac{az + b}{cz + d}, \quad ad - bc = 1, \tag{4.13}$$

determines a motion of the Lobachevsky plane (in Poincaré's model) if

$$\frac{dw\, d\bar{w}}{(1 - |w|^2)^2} = \frac{dz\, d\bar{z}}{(1 - |z|^2)^2}, \tag{4.14}$$

and it maps the circle $|z| < 1$ into the circle $|w| < 1$. We write (4.14) in more detail:

$$\begin{aligned}\frac{dw\, d\bar{w}}{(1 - |w|^2)^2} &= \frac{dz\, d\bar{z}}{(|cz + d|^2 - |az + b|^2)} \\ &= \frac{dz\, d\bar{z}}{((|c|^2 - |a|^2) z\bar{z} + (c\bar{d} - a\bar{b})z + (\bar{c}d - \bar{a}b)\bar{z} + (|d|^2 - |b|^2))^2}.\end{aligned}$$

This implies that equation (4.14) holds whenever

$$|a|^2 - |c|^2 = |d|^2 - |b|^2 = \pm 1, \qquad a\bar{b} - c\bar{d} = 0.$$

Condition $|w| < 1$ implies the equalities

$$|a|^2 - |c|^2 = 1, \qquad |d|^2 - |b|^2 = -1.$$

Therefore, the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{C})$ of a linear-fractional transformation which is a motion of the Lobachevsky plane belongs to the group $\mathrm{SU}(1, 1)$.

The proper transformations (orthochrone and preserving orientation) in $\mathrm{O}(1, 2)$ form a group isomorphic to $\mathrm{SU}(1, 1)/\pm 1$.

If the linear-fractional transformation (4.13) maps the upper half-plane $y > 0$ into itself, then $a$, $b$, $c$, and $d$ are real numbers. Such a transformation is an orientation-preserving motion of the half-plane with hyperbolic metric. The calculations justifying this statement and the fact that such transformations determine all the proper motions are similar to the previous arguments, and we omit them. These transformations form the group $\mathrm{SL}(2, \mathbb{R})/\pm 1$.

Thus we have deduced the following corollary.

**Corollary 4.2.** *The groups* $\mathrm{SU}(1,1)/\pm 1$, $\mathrm{SL}(2)/\pm 1$ *and the subgroup* $\mathrm{O}(1,2)$ *formed by proper transformations are isomorphic to each other.*

A detailed proof of this corollary repeats almost verbatim the proof of Theorem 4.4 with Lemma 4.9 used instead of Lemma 4.7.

Each of these groups determines all the proper motions of the Lobachevsky plane in one of the three models (Poincaré, half-plane, pseudosphere). The complete groups of motions are obtained by supplementing these groups with the motion that reverses orientation and with all its compositions with the proper motions. In the half-plane model this motion is the mapping $z \to -\bar{z}$, and in Poincaré's model, it is the conjugation $z \to \bar{z}$.

**4.3.4. Hilbert's theorem on impossibility of imbedding the pseudosphere into $\mathbb{R}^3$.** In Section 3.4.2 we have shown that the Codazzi equations for a surface of constant negative curvature reduce to the sine-Gordon equation

$$\omega_{xy} + \sin\omega = 0,$$

where $x, y$ are the asymptotic coordinates on the plane in which the first and the second fundamental forms are

$$g_{ij}\, dx^i\, dx^j = dx^2 + 2\cos\omega\, dx\, dy + dy^2, \quad b_{ij}\, dx^i\, dx^j = 2\sin\omega\, dx\, dy, \tag{4.15}$$

where $x^1 = x, x^2 = y$, and $\omega$ is the angle between the asymptotic directions. Therefore, if we have a solution $\omega$ of the sine-Gordon equation which satisfies the geometric condition

$$0 < \omega < \pi$$

and is defined in some domain $U$, then this solution enables us to construct an imbedding of $U \subset \mathbb{R}^2$ into $\mathbb{R}^3$ for which the first and the second fundamental forms are given by (4.15).

As was shown by Hilbert, there is no such solution defined on the entire plane. This implies the following theorem.

**Theorem 4.6.** *The Lobachevsky plane cannot be imbedded into* $\mathbb{R}^3$.

This theorem says that there is no metric imbedding for which the metric induced on the two-dimensional plane is the Lobachevsky metric.

**Proof.** We present an outline of the proof using the fact that if such an imbedding existed, then the asymptotic coordinates could be introduced globally. We do not prove this fact, which follows from the theory of ordinary differential equations.

Suppose that such an imbedding does exist. Its area form is

$$\sqrt{g}\, dx\, dy = \sin\omega\, dx\, dy.$$

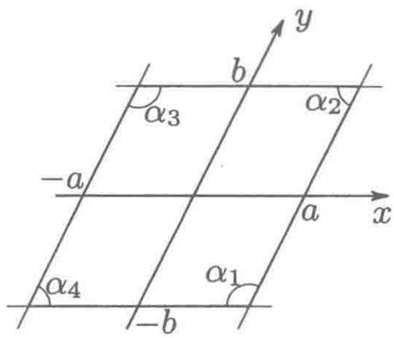


**Figure 4.4.** A "parallelogram" on the Lobachevsky plane.

Consider a "parallelogram" $\Pi_{a,b}$ defined by the conditions $-a \le x \le a$ and $-b \le y \le b$ (see Figure 4.4). Its area is $\int_{-a}^{a}\int_{-b}^{b} \sin\omega\,dx\,dy$, and according to the sine-Gordon equation this integral is equal to

$$\int_{-a}^{a}\int_{-b}^{b} \sin\omega\,dx\,dy = -\int_{-a}^{a}\int_{-b}^{b} \omega_{xy}\,dx\,dy = \alpha_1+\alpha_2+\alpha_3+\alpha_4-2\pi,$$

where the $\alpha_i$ are the inner angles of the "parallelogram" $\Pi_{a,b}$. We have $0<\alpha_i<\pi$, $i=1,\dots,4$, which implies that the area of $\Pi_{a,b}$ is bounded from above by $2\pi$ for all $a,b$. As $a,b\to\infty$, these "parallelograms" asymptotically cover the entire plane, whereas we know that the area of the pseudosphere is infinite. Thus we arrived at a contradiction, which proves the theorem. □

## 4.4. The theory of surfaces in terms of a conformal parameter

**4.4.1. Existence of a conformal parameter.** Consider a smooth parametrized surface $r=r(u,v)$ in $\mathbb{R}^3$. Let

$$dl^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$$

be its first fundamental form.

**Theorem 4.7.** *In a neighborhood of any point of a smooth surface one can introduce conformal coordinates $x$ and $y$, i.e., the coordinates such that*

$$dl^2 = g(x,y)(dx^2+dy^2).$$

For the proof we need some auxiliary results. First, we derive equations for the transition functions to the new coordinates,

$$(u,v)\to(x(u,v),y(u,v)).$$

**Lemma 4.10.** *If $x$, $y$ are conformal coordinates, then*

$$\begin{cases} \dfrac{\partial x}{\partial u} = \dfrac{1}{\sqrt{EG-F^2}}\Big(E\,\dfrac{\partial y}{\partial v} - F\,\dfrac{\partial y}{\partial u}\Big), \\ \dfrac{\partial x}{\partial v} = \dfrac{1}{\sqrt{EG-F^2}}\Big(F\,\dfrac{\partial y}{\partial v} - G\,\dfrac{\partial y}{\partial u}\Big), \end{cases} \tag{4.16}$$

$$\begin{cases} \dfrac{\partial y}{\partial u} = \dfrac{1}{\sqrt{EG-F^2}}\Big(-E\,\dfrac{\partial x}{\partial v} + F\,\dfrac{\partial x}{\partial u}\Big), \\ \dfrac{\partial y}{\partial v} = \dfrac{1}{\sqrt{EG-F^2}}\Big(-F\,\dfrac{\partial x}{\partial v} + G\,\dfrac{\partial x}{\partial u}\Big). \end{cases} \tag{4.17}$$

*The systems* (4.16) *and* (4.17) *are equivalent.*

**Proof.** The equality

$$g(dx^2+dy^2) = E\,du^2 + 2F\,du\,dv + G\,dv^2$$

can be transformed as follows:

$$g\Big(\Big(\frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv\Big)^2 + \Big(\frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial v}\,dv\Big)^2\Big) = E\,du^2 + 2F\,du\,dv + G\,dv^2.$$

Therefore, it is equivalent to the system

$$\begin{aligned} \Big(\frac{\partial x}{\partial u}\Big)^2 + \Big(\frac{\partial y}{\partial u}\Big)^2 &= \frac{1}{g}E, \\ \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} &= \frac{1}{g}F, \\ \Big(\frac{\partial x}{\partial v}\Big)^2 + \Big(\frac{\partial y}{\partial v}\Big)^2 &= \frac{1}{g}G, \end{aligned}$$

which can be rewritten as two linear systems

$$A\begin{pmatrix} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial u} \end{pmatrix} = \frac{1}{g}\begin{pmatrix} E \\ F \end{pmatrix}, \qquad A\begin{pmatrix} \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v} \end{pmatrix} = \frac{1}{g}\begin{pmatrix} F \\ G \end{pmatrix}, \tag{4.18}$$

where $A$ is the Jacobi matrix of the change of coordinates: $A = \left(\begin{smallmatrix} x_u & y_u \\ x_v & y_v \end{smallmatrix}\right)$. Actually we only restated the transition formulas for the Gram matrix:

$$A^\top \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} A = \begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

Now we extract the square roots from the determinants of the left- and right-hand sides of this equality to obtain

$$g = \frac{\sqrt{EG-F^2}}{\det A}.$$

Substituting this expression for $g$ into equations (4.18) and solving these systems formally we obtain the systems (4.16) and (4.17). □

To every metric $g_{jk}\,dx^j\,dx^k$ corresponds the *Laplace–Beltrami operator* acting on a function $\varphi$ (on the surface) by the formula

$$\Delta\varphi = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^j}\Big(\sqrt{g}g^{jk}\frac{\partial\varphi}{\partial x^k}\Big).$$

Solutions to the *Beltrami equation*

$$\Delta\varphi = 0$$

are called *harmonic functions* (on the surface). For the Euclidean metric $g_{jk} = \delta_{jk}$ the Laplace–Beltrami operator becomes the Laplace operator

$$\Delta = \frac{\partial^2}{(\partial x^1)^2} + \cdots + \frac{\partial^2}{(\partial x^n)^2},$$

and in what follows, by $\Delta$ we mean this operator unless otherwise stated.

**Lemma 4.11.** *If smooth functions $x(u,v)$ and $y(u,v)$ satisfy equations* (4.16) *and* (4.17), *then they are harmonic:*

$$\Delta x = \Delta y = 0.$$

**Proof.** Substitute relations (4.16) into the obvious equation

$$\frac{\partial^2 x}{\partial u \partial v} - \frac{\partial^2 x}{\partial v \partial u} = 0$$

to obtain

$$\frac{\partial}{\partial u}\Big(\frac{1}{\sqrt{EG-F^2}}\Big(F\frac{\partial y}{\partial v} - G\frac{\partial y}{\partial u}\Big)\Big) - \frac{\partial}{\partial v}\Big(\frac{1}{\sqrt{EG-F^2}}\Big(E\frac{\partial y}{\partial v} - F\frac{\partial y}{\partial u}\Big)\Big) = 0.$$

Setting $E = g_{11}$, $F = g_{12}$, and $G = g_{22}$, this equation can be rewritten as $\Delta y = 0$. In a similar way we prove that $\Delta x = 0$. □

Next, we show by a direct substitution that if $x(u,v)$ and $y(u,v)$ satisfy equations (4.16) and (4.17), then the Jacobian of the mapping $(u,v) \to (x,y)$ equals

$$\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} = \frac{1}{\sqrt{EG-F^2}}\langle V, V\rangle, \tag{4.19}$$

where $V = (\partial x/\partial v, -\partial x/\partial u)$ and the scalar product is specified by the metric of the surface. Indeed,

$$\begin{aligned}
&\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \\
&\quad = \frac{\partial x}{\partial u}\frac{1}{\sqrt{EG-F^2}}\Big(-F\frac{\partial x}{\partial v} + G\frac{\partial x}{\partial u}\Big) - \frac{\partial x}{\partial v}\frac{1}{\sqrt{EG-F^2}}\Big(-E\frac{\partial x}{\partial v} + F\frac{\partial x}{\partial u}\Big) \\
&\quad = \frac{1}{\sqrt{EG-F^2}}\Big(G\Big(\frac{\partial x}{\partial u}\Big)^2 - 2F\frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + E\Big(\frac{\partial x}{\partial v}\Big)^2\Big).
\end{aligned}$$

For the proof of Theorem 4.7 we also need the following lemma.

**Lemma 4.12.** *In a neighborhood of any point* $(u_0, v_0)$ *of the surface there exist*

1) *a harmonic function* $x(u,v)$ *such that in this neighborhood*

$$\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial x}{\partial v}\right)^2 \neq 0; \tag{4.20}$$

2) *a function* $y(u,v)$ *satisfying equations* (4.17).

We state this lemma without proof. Note only that when the mapping $r(u,v)$ is analytic, the solutions $x(u,v)$ and $y(u,v)$ can be constructed in the form of convergent series.

**Proof of Theorem 4.7.** Take a harmonic function $x(u,v)$ satisfying condition (4.20) in a neighborhood of a point and, based on this function, construct a solution $y(u,v)$ to equations (4.17). Formulas (4.19) and (4.20) imply that the Jacobian of the mapping $(u,v) \to (x,y)$ does not vanish anywhere. Hence, by the inverse function theorem, the functions $x$ and $y$ specify local coordinates in a small neighborhood of the point $(u_0, v_0)$. Formulas (4.16) and (4.17) imply that these coordinates are conformal, which completes the proof. □

**4.4.2. The basic equations in terms of a conformal parameter.** Let $x^1 = x$, $x^2 = y$ be conformal coordinates:

$$g_{jk}\, dx^j\, dx^k = g(dx^2 + dy^2).$$

Then substituting $g_{jk}$ into the formula

$$\Gamma^i_{jk} = \frac{1}{2} g^{il}\left(\frac{\partial g_{jl}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l}\right)$$

we find the Christoffel symbols:

$$\begin{aligned} \Gamma^1_{11} = \Gamma^2_{12} = \Gamma^2_{21} = -\Gamma^1_{22} = \frac{1}{2}\frac{\partial \log g}{\partial x^1}, \\ \Gamma^2_{22} = \Gamma^1_{12} = \Gamma^1_{21} = -\Gamma^2_{11} = \frac{1}{2}\frac{\partial \log g}{\partial x^2}. \end{aligned} \tag{4.21}$$

By Theorem 3.6 we obtain the formula for the Gaussian curvature,

$$K = -\frac{1}{2g}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\log g.$$

Another way to obtain this formula is to derive it as a consequence of the Codazzi equations written in terms of the conformal parameter

$$z = x + iy.$$

Denote by $r_z$ and $r_{\bar{z}}$ the derivatives of $r$ with respect to $z$ and $\bar{z}$. Extend the scalar product in $\mathbb{R}^3$,

$$\langle \xi, \eta \rangle = \xi^1\eta^1 + \xi^2\eta^2 + \xi^3\eta^3,$$

to complex vectors $\xi = \xi^j e_j$ and $\eta = \eta^k e_k$, where $\xi^j$, $\eta^k \in \mathbb{C}$, $j, k = 1, 2, 3$.

**Lemma 4.13.** *The parameter $z = x + iy$ on a surface $r = r(x, y)$ is conformal if and only if*

$$\langle r_z, r_z \rangle = 0.$$

**Proof.** The left-hand side can be written as

$$\frac{1}{4}(E - G) - \frac{i}{2} F,$$

and it vanishes if and only if $E = G$, $F = 0$. □

Using formulas (4.21), rewrite the derivational equations in the form

$$\frac{\partial}{\partial z}\begin{pmatrix} r_z \\ r_{\bar{z}} \\ n \end{pmatrix} = U \begin{pmatrix} r_z \\ r_{\bar{z}} \\ n \end{pmatrix}, \qquad \frac{\partial}{\partial \bar{z}}\begin{pmatrix} r_z \\ r_{\bar{z}} \\ n \end{pmatrix} = V \begin{pmatrix} r_z \\ r_{\bar{z}} \\ n \end{pmatrix}, \tag{4.22}$$

where

$$U = \begin{pmatrix} \frac{\partial}{\partial z}\log g & 0 & A \\ 0 & 0 & B \\ -\frac{2B}{g} & -\frac{2A}{g} & 0 \end{pmatrix}, \qquad V = \begin{pmatrix} 0 & 0 & B \\ 0 & \frac{\partial}{\partial \bar{z}}\log g & \bar{A} \\ -\frac{2\bar{A}}{g} & -\frac{2B}{g} & 0 \end{pmatrix},$$

$$A = \langle r_{zz}, n \rangle, \qquad B = \langle r_{z\bar{z}}, n \rangle,$$

with $n$ being the normal vector to the surface.

Let us write down the second fundamental form

$$b_{jk}\, dx^j\, dx^k = \langle r_{jk}, n \rangle\, dx^j\, dx^k$$

in terms of $A$ and $B$. Since

$$r_{11} = \frac{\partial^2 r}{\partial x^2} = \Big(\frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}}\Big)^2 r = r_{zz} + 2r_{z\bar{z}} + r_{\bar{z}\bar{z}},$$

$$r_{12} = \frac{\partial^2 r}{\partial x \partial y} = i\Big(\frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}}\Big)\Big(\frac{\partial}{\partial z} - \frac{\partial}{\partial \bar{z}}\Big) r = i(r_{zz} - r_{\bar{z}\bar{z}}),$$

$$r_{22} = \frac{\partial^2 r}{\partial y^2} = \Big[i\Big(\frac{\partial}{\partial z} - \frac{\partial}{\partial \bar{z}}\Big)\Big]^2 r = -r_{zz} + 2r_{z\bar{z}} - r_{\bar{z}\bar{z}},$$

we have

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} 2B + A + \bar{A} & i(A - \bar{A}) \\ i(A - \bar{A}) & 2B - (A + \bar{A}) \end{pmatrix}.$$

Hence we immediately obtain the formulas for the mean and Gaussian curvature:

$$H = \frac{2B}{g}, \qquad K = \frac{4(B^2 - |A|^2)}{g^2}. \tag{4.23}$$

The Codazzi equations provide a condition for the derivational equations (4.22) to be compatible. They have the form

$$\frac{\partial U}{\partial \bar{z}} - \frac{\partial V}{\partial z} + [U, V] = 0.$$

One can check by a direct calculation that they reduce to two equations:

$$\frac{\partial^2}{\partial z \partial \bar{z}} \log g + \frac{2}{g}(B^2 - |A|^2) = 0, \qquad \frac{\partial A}{\partial \bar{z}} - \frac{\partial B}{\partial z} + \frac{\partial \log g}{\partial z} B = 0.$$

Substituting the formulas (4.23) for the curvatures into these equations, we reduce the Codazzi equations to the form

$$K = -\frac{2}{g} \frac{\partial^2}{\partial z \partial \bar{z}} \log g, \qquad \frac{\partial A}{\partial \bar{z}} = \frac{g}{2} \frac{\partial H}{\partial z}. \tag{4.24}$$

The latter equation falls into two real equations for the real and imaginary parts of the expressions. In this way we obtain three real Codazzi equations (see Section 3.4.2).

**4.4.3. Hopf differential and its applications.** The expression $A\,dz^2$, where $A = \langle r_{zz}, n \rangle$, is called the *Hopf differential* of the surface. Under a complex-analytic change of coordinates $z \to w$ (when changing to another conformal parameter) it transforms by the rule

$$A\,dz^2 \to \widetilde{A}\,dw^2 = \widetilde{A}\Big(\frac{\partial w}{\partial z}\Big)^2 dz^2,$$

i.e.,

$$A = \widetilde{A}\Big(\frac{\partial w}{\partial z}\Big)^2. \tag{4.25}$$

Differentials $A\,dz^2 = \widetilde{A}\,dw^2$ satisfying (4.25) are called *quadratic differentials.*

Analytic properties of the Hopf differential are related to certain geometric properties of the surface.

**Theorem 4.8.** 1. *The quadratic Hopf differential $A\,dz^2$ vanishes at a point ($A = 0$) if and only if the principal curvatures at this point are equal to each other: $k_1 = k_2$.*

2. *The Hopf differential $A\,dz^2$ is holomorphic (i.e., $\partial A/\partial \bar{z} = 0$) in some domain if and only if the mean curvature takes a constant value in this domain: $H \equiv$ const.*

**Proof.** The principal curvatures are equal if and only if the discriminant of the equation

$$p(\lambda) = \det \begin{pmatrix} L - \lambda E & M - \lambda F \\ M - \lambda F & N - \lambda G \end{pmatrix} = 0$$

is equal to zero. But in our case the discriminant vanishes exactly when

$$(A + \bar{A})^2 - (A - \bar{A})^2 = 4(\operatorname{Re} A)^2 + 4(\operatorname{Im} A)^2 = 0,$$

which is equivalent to $A = 0$.

The equality $\partial H/\partial z = 0$ can be rewritten as

$$\frac{\partial H}{\partial z} = \frac{1}{2}\Big(\frac{\partial H}{\partial x} - i\,\frac{\partial H}{\partial y}\Big),$$

and since $H$ is a real function, this equality is equivalent to $\operatorname{grad} H = 0$. Now the theorem follows from the Codazzi equations (4.24). □

A surface point where the principal curvatures are equal to each other is said to be *umbilical*.

The following assertion known as the Darboux theorem is deduced from Theorem 4.8.

**Theorem 4.9.** *If all the points of a domain on a surface are umbilical, then this domain lies either on a plane in* $\mathbb{R}^3$ *or on a sphere of radius* $R$.

**Proof.** Condition $A \equiv 0$ implies that $H = \text{const}$. Hence, integrating the derivational equation $n_z = -Hr_z$ with respect to $z$, we obtain

$$n = -Hr + r_0.$$

If $H = 0$, then the normal $n$ is constant, and the surface coincides with a domain on the plane orthogonal to $n$. If $H \neq 0$, we shift the surface by $r_0/H$: $r \to \tilde{r} = r - r_0/H$. Then we have

$$\langle H\tilde{r}, H\tilde{r}\rangle = \langle n, n\rangle = 1$$

and $\langle \tilde{r}, \tilde{r}\rangle = 1/H^2$. Hence the shifted surface lies on the sphere of radius $1/H$. □

New we present some applications of the Codazzi equations in complex form.

**4.4.4. Surfaces of constant Gaussian curvature. The Liouville equation.** Let $g(z, \bar{z})\, dz\, d\bar{z}$ be a metric on a surface and let $g = e^{\varphi}$. The Gaussian curvature equals

$$K = -2e^{-\varphi}\frac{\partial^2 \varphi}{\partial z \partial \bar{z}}. \tag{4.26}$$

If it is constant, then the function $\varphi$ satisfies the *Liouville equation*

$$\Delta\varphi = -2Ke^{\varphi}.$$

Like the sine-Gordon equation, this equation is also used in the theory of solitons; it is integrable by simpler methods. The general solution was found by Liouville and, say, for $K \equiv -1$ it has the form

$$\varphi(x,y) = \log\Big[\frac{4A'(z)B'(\bar{z})}{(A(z)+B(\bar{z}))^2}\Big],$$

where $A(\omega)$, $B(\omega)$ are arbitrary complex-analytic functions and $z = x+iy$. From this formula one can derive general solutions for other $K \neq 0$.

**Theorem 4.10.** *If the Gaussian curvature is constant, then the surface is locally isometric to*:

1) *a sphere for* $K > 0$;
2) *a plane for* $K = 0$;
3) *a pseudosphere for* $K < 0$.

**Proof.** The equality (4.26) implies that

$$0 = \frac{\partial}{\partial z}\Big(-\frac{K}{2}\Big) = \frac{\partial}{\partial z}\Big(e^{-\varphi}\frac{\partial^2\varphi}{\partial z\partial\bar{z}}\Big)$$
$$= e^{-\varphi}\Big(\frac{\partial^3\varphi}{\partial z^2\partial\bar{z}} - \frac{\partial\varphi}{\partial z}\frac{\partial^2\varphi}{\partial z\partial\bar{z}}\Big) = e^{-\varphi}\frac{\partial}{\partial\bar{z}}\Big(\frac{\partial^2\varphi}{\partial z^2} - \frac{1}{2}\Big(\frac{\partial\varphi}{\partial z}\Big)^2\Big).$$

Thus the function

$$\psi(z) = \frac{\partial^2\varphi}{\partial z^2} - \frac{1}{2}\Big(\frac{\partial\varphi}{\partial z}\Big)^2$$

is holomorphic.

Find a conformal parameter $w$ on the surface such that

$$\widetilde{\psi}(w) = \frac{\partial^2\widetilde{\varphi}}{\partial w^2} - \frac{1}{2}\Big(\frac{\partial\widetilde{\varphi}}{\partial w}\Big)^2 = 0.$$

To this end observe that the change of coordinates $z = f(w)$ transforms the metric by the formulas

$$e^{\widetilde{\varphi}(w,\bar{w})} = e^{\varphi(z,\bar{z})}\Big|\frac{df}{dw}\Big|^2, \qquad \widetilde{\varphi}(w,\bar{w}) = \varphi(z,\bar{z}) + \log\frac{df}{dw} + \log\frac{d\bar{f}}{d\bar{w}}.$$

The function $\widetilde{\psi}(w)$ equals

$$\widetilde{\psi}(w) = \frac{\partial^2\widetilde{\varphi}}{\partial w^2} - \frac{1}{2}\Big(\frac{\partial\widetilde{\varphi}}{\partial w}\Big)^2 = \psi(z)(f')^2 + \frac{f'''}{f'} - \frac{3}{2}\Big(\frac{f''}{f'}\Big)^2,$$

where $f' = \frac{df}{dw}$. The expression

$$\frac{f'''}{f'} - \frac{3}{2}\Big(\frac{f''}{f'}\Big)^2$$

is called the *Schwarz derivative*. It can be shown that the equation

$$\frac{f'''}{f'} - \frac{3}{2}\Big(\frac{f''}{f'}\Big)^2 = -\psi(f(w))(f')^2 \tag{4.27}$$

is locally solvable for $f$. Indeed, the function $\psi(z)$ is holomorphic and representable in a neighborhood of any point by a convergent series in powers of $z$. Substituting this series into equation (4.27), one can find a local solution in the form of a convergent series in powers of $w$.

We will use the solution to this equation for the change of coordinates. We obtain

$$\frac{\partial^2 e^{-\widetilde{\varphi}/2}}{\partial w^2} = -\frac{1}{2}\, e^{-\widetilde{\varphi}/2}\Big(\frac{\partial^2\widetilde{\varphi}}{\partial w^2} - \frac{1}{2}\Big(\frac{\partial\widetilde{\varphi}}{\partial w}\Big)^2\Big) = 0.$$

Since $e^{-\widetilde{\varphi}/2}$ is a real function, we have

$$\frac{\partial^2 e^{-\widetilde{\varphi}/2}}{\partial \bar{w}^2} = 0.$$

This implies that

$$e^{-\widetilde{\varphi}/2} = aw\bar{w} + bw + \bar{b}\bar{w} + c,$$

where $a$, $c$ are real constants and $b$ is a complex constant. The metric of the surface becomes

$$g(z,\bar{z})\, dz\, d\bar{z} = \frac{dw\, d\bar{w}}{(aw\bar{w} + bw + \bar{b}\bar{w} + c)^2}, \tag{4.28}$$

and its curvature is equal to

$$K = 4(ac - b\bar{b}).$$

The form (4.28) is reduced by linear-fractional transformations to one of the following forms:

$$\begin{aligned} \frac{4R^2\, dz\, d\bar{z}}{(1+|z|^2)^2} \quad &\text{for} \quad K = \frac{1}{R^2} > 0, \\ dz\, d\bar{z} \quad &\text{for} \quad K = 0, \\ \frac{4R^2\, dz\, d\bar{z}}{(1-|z|^2)^2} \quad &\text{for} \quad K = -\frac{1}{R^2} < 0. \end{aligned}$$

These are metrics of the sphere, plane, and pseudosphere. □

**4.4.5. Surfaces of constant mean curvature. The sinh-Gordon equation.** It is not hard to show that the holomorphic differential on a sphere is identically equal to zero. Hence Theorem 4.9 implies that if a sphere imbedded into $\mathbb{R}^3$ has a constant mean curvature $H$, then it is specified by the equation

$$x^2 + y^2 + z^2 = \frac{1}{H^2}$$

(in appropriate Euclidean coordinates). The imbedding of a torus is specified by a doubly periodic mapping into $\mathbb{R}^3$:

$$r \colon \mathbb{R}^2 \to \mathbb{R}^3, \quad r(x+\lambda) = r(x),$$

where $\lambda \in \Lambda$ and $\Lambda$ is a lattice in $\mathbb{R}^2$, i.e., the set of vectors of the form $p\tilde{e}_1 + q\tilde{e}_2$, where the vectors $\tilde{e}_1$, $\tilde{e}_2$ are linearly independent and $p, q \in \mathbb{Z}$.

The uniformization theorem, which is a deep and important fact of complex analysis, states, in particular, that the conformal parameter on the imbedded torus can be chosen globally. This means that the torus is specified by an imbedding

$$r \colon \mathbb{C} \to \mathbb{R}^3,$$

which is periodic relative to a lattice $\widetilde{\Lambda}$ (possibly different from $\Lambda$),

$$r(z+\tilde{\lambda}) = r(z), \quad \lambda \in \widetilde{\Lambda},$$

and the first fundamental form is

$$e^{\varphi(z,\bar{z})}\, dz\, d\bar{z}.$$

The lattice $\widetilde{\Lambda}$ is uniquely defined up to affine transformations of the complex plane $z \to \alpha z + \beta$ and determines the *conformal class* of a metric on the torus. Therefore, we take a global conformal parameter $z$ on the torus.

The Hopf differential on a torus has the form $A\, dz^2 = f(z,\bar{z})\, dz^2$, where the function $f$ is doubly periodic relative to shifts by $\lambda \in \Lambda$. If $H = \text{const}$, then $f$ is a holomorphic function. But any doubly periodic holomorphic function is a constant, hence $A = \text{const}$. If $A = 0$, then all the points are umbilic, and by Theorem 4.9 the imbedded torus must lie either on a plane, or on a sphere, which is, obviously, impossible. Set $\varphi = \log g$ and rewrite the first of the Codazzi equations in the form

$$\partial\bar{\partial}\varphi + 2e^{-\varphi}(B^2 - |A|^2) = 0. \tag{4.29}$$

A homothety $r \to \lambda r$ changes the mean curvature by the formula $H \to H\lambda^{-1}$; hence we can take $H = 1/2$. In a similar way, changing the conformal parameter $z$ to $\mu z$ with $\mu = \text{const}$, we can reduce $A$ to $A = 1/4$. Now, substituting the expressions for $A$ and $B = He^{\varphi}/2$ into (4.29), we obtain the equation

$$\Delta\varphi + \sinh\varphi = 0.$$

Like the sine-Gordon equation, this equation belongs to the class of integrable (soliton) equations. However, it differs from the former by the following two features.

1. In contrast to the sine-Gordon equation, this equation is elliptic. Consequently, every doubly periodic solution of it without singularities is determined by a Riemann surface of finite type. This property does not hold in the hyperbolic case typical for the theory of solitons; it is only

known that similar so-called "finite-zone" solutions are everywhere dense in some cases. There are also various classes of Riemann surfaces of infinite type, which have been explored only in some particular settings.

2. Based on any doubly periodic smooth solution of this equation, one can construct an imbedding of the complex plane $\mathbb{C}$ into $\mathbb{R}^3$ such that the mean curvature is constant, $H = \frac{1}{2}$, and the bases $(r_x, r_y, n)$ are periodic relative to the lattice $\widetilde{\Lambda}$. However the resulting surface need not roll up into a torus, but it may remain an imbedded plane or roll up into a cylinder. It turned out that among such surfaces, indeed, there exist tori of constant mean curvature, but to describe the situations that give rise to a torus is a separate problem, which is solved in terms of analysis on Riemann surfaces (see [**6, 53**]).

## 4.5. Minimal surfaces

**4.5.1. The Weierstrass–Enneper formulas for minimal surfaces.** There is an elegant application of the conformal parameter that provides a general method for constructing minimal surfaces from a pair of complex-analytic functions. We present this method in this section.

The derivational equations (4.22) imply that

$$\frac{\partial^2 r}{\partial z \partial \bar{z}} = \frac{1}{2} H g n, \tag{4.30}$$

where $n$ is the normal vector and $g\, dz\, d\bar{z}$ is a metric on the surface $r(z, \bar{z})$.

The surface is said to be *minimal* if its mean curvature vanishes everywhere, $H \equiv 0$. Formula (4.30) implies that a minimal surface satisfies the equality

$$\frac{\partial^2 r}{\partial z \partial \bar{z}} = 0,$$

i.e., the coordinate functions $x^j(z, \bar{z})$, $j = 1, 2, 3$, are harmonic relative to the conformal coordinates:

$$\Delta x^j = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) x^j = 0,$$

where $z = x + iy$. This property can be taken for the definition of a minimal surface.

We see that the functions

$$\frac{\partial x^j}{\partial z}, \quad j = 1, 2, 3,$$

are complex-analytic. Since the parameter $z$ is conformal, Lemma 4.13 implies that

$$\sum_{j=1}^{3}\left(\frac{\partial x^j}{\partial z}\right)^2 \equiv 0.$$

Since the surface is regular and its metric is

$$2\left(\sum_{j=1}^{3}\left|\frac{\partial x^j}{\partial z}\right|^2\right) dz\, d\bar{z},$$

the vector-function $(\partial x^1/\partial z, \partial x^2/\partial z, \partial x^3/\partial z)$ has no zeros.

It turns out that from any vector-function $\varphi = (\varphi_1, \varphi_2, \varphi_3)$ satisfying these conditions one can locally construct a minimal surface. And what is more, this is true for any such vector-function defined in a simply connected domain on $\mathbb{C}$.

A domain $U \subset \mathbb{C}$ is said to be *simply connected* if each continuous mapping $f\colon S^1 \to U$ of the unit circle $\{|z| = 1\} \subset \mathbb{C}$ into $U$ can be extended to a continuous mapping of the entire unit disk $\{|z| \leq 1\}$ into $U$. The disks $\{|z| \leq R\}$ provide immediate examples of simply connected domains. The Stokes theorem to be proved in Chapter 9 implies that for any complex-analytic function $f(z)$ in a simply connected domain $U$ the integrals $\int f(z)\, dz$ and $\int \overline{f(z)}\, d\bar{z}$ are completely determined by the initial and terminal points of the integration path. In the forthcoming theorem this fact will be needed to make sure that the notion of minimal surface is well defined.

**Theorem 4.11.** *Let $U$ be a simply connected domain on the complex plane $\mathbb{C}$. Let*

$$\varphi\colon U \to \mathbb{C}^3$$

*be a vector-function defined in this domain and satisfying the following conditions:*

1) *$\varphi$ is a complex-analytic function (i.e., each of its components $\varphi_1$, $\varphi_2$, and $\varphi_3$ is complex-analytic);*

2) *$\varphi$ has no zeros in $U$;*

3) $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 \equiv 0$.

*Then there exists a regular minimal surface $r(z, \bar{z})$ such that*

$$\varphi = \frac{\partial r}{\partial z} = \left(\frac{\partial x^1}{\partial z}, \frac{\partial x^2}{\partial z}, \frac{\partial x^3}{\partial z}\right),$$

*and $z$ is a conformal parameter on this surface.*

**Proof.** We know from the theory of analytic functions that if $f$ is complex-analytic, then the value of the integral

$$\int_{z_0}^{z} (f(w)\, dw + \overline{f(w)}\, d\bar{w})$$

does not depend on the choice of the integration path joining the points $z_0$ and $z$ in the simply connected domain $U$. Therefore, the functions

$$x^j(z, \bar{z}) = \int_{z_0}^{z} (\varphi_j(w)\, dw + \overline{\varphi_j(w)}\, d\bar{w}) \tag{4.31}$$

are well defined. They specify a mapping

$$r \colon U \to \mathbb{R}^3, \quad r = (x^1, x^2, x^3),$$

and it is obvious that

$$\frac{\partial r}{\partial z} = \left(\frac{\partial x^1}{\partial z}, \frac{\partial x^2}{\partial z}, \frac{\partial x^3}{\partial z}\right) = (\varphi_1, \varphi_2, \varphi_3).$$

Since the vector-function $\varphi$ has no zeros, we have $\langle r_z, r_{\bar{z}}\rangle \neq 0$ everywhere. But

$$\langle r_z, r_{\bar{z}}\rangle = \frac{1}{4}\left(\langle r_x, r_x\rangle + \langle r_y, r_y\rangle\right)$$

and

$$\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = \langle r_z, r_z\rangle = \frac{1}{4}\left(\langle r_x, r_x\rangle - \langle r_y, r_y\rangle\right) - \frac{i}{2}\langle r_x, r_y\rangle = 0.$$

This implies that the functions $r_x$ and $r_y$ are linearly independent everywhere and the parameter $z$ is conformal.

It remains to note that $r_{z\bar{z}} = 0$. Relation (4.30) implies that the mean curvature is equal to zero everywhere, i.e., the surface is minimal. □

The general solution of the equation

$$\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$$

has the form

$$\varphi_1 = \frac{1}{2}(\chi_1^2 - \chi_2^2), \qquad \varphi_2 = \frac{i}{2}(\chi_1^2 + \chi_2^2), \qquad \varphi_3 = \chi_1\chi_2, \tag{4.32}$$

and two such pairs, $(\chi_1, \chi_2)$ and $(-\chi_1, -\chi_2)$, correspond to any nonzero solution. Therefore, if we are given two complex-analytic functions $\chi_1$ and $\chi_2$ that do not vanish simultaneously in a simply connected domain $U$, then these formulas enable us to construct the vector-function $\varphi$. This function satisfies all the conditions of Theorem 4.11, and using this function we can construct the minimal surface. The metric on this surface is

$$2\langle r_z, r_{\bar{z}}\rangle\, dz\, d\bar{z} = (|\chi_1|^2 + |\chi_2|^2)^2\, dz\, d\bar{z}. \tag{4.33}$$

Usually the formulas specifying a surface by a pair of complex-analytic functions are written in terms of functions $f$ and $g$ that satisfy the relations

$$f = \chi_1^2, \qquad g^2 f = \chi_2^2. \tag{4.34}$$

Substituting the equalities (4.32) and (4.34) into (4.31), we obtain the following *Weierstrass–Enneper formulas*:

$$\begin{aligned} x^1(z,\bar z) &= \frac{1}{2}\int ((1-g^2)f\,dw + \overline{(1-g^2)f}\,d\bar w),\\ x^2(z,\bar z) &= \frac{i}{2}\int ((1+g^2)f\,dw - \overline{(1+g^2)f}\,d\bar w),\\ x^3(z,\bar z) &= \int (fg\,dw + \overline{fg}\,d\bar w). \end{aligned} \tag{4.35}$$

The integration is along any path from $z_0$ to $z$ in the domain $U$. Obviously, we can add any constants specifying $r(z_0)$ in the right-hand side. It is clear from the construction that these formulas locally give any minimal surface.

### 4.5.2. Examples of minimal surfaces.

EXAMPLE 1. A PLANE. In this case $f = z^2$, $g = 1$ (or $\chi_1 = \chi_2 = z$ and $z \in \mathbb{C}$).

A plane may be specified as the graph of a function $z = f(x, y)$ by taking for $f$ a trivial function $f = \text{const}$ on the entire plane $xy$.

In general, a surface that locally is the graph of a function $z = f(x, y)$ is minimal if

$$H = \frac{1}{2}\,\frac{(1+f_y^2)f_{xx} + (1+f_x^2)f_{yy} - 2f_xf_yf_{xy}}{1+f_x^2+f_y^2} = 0$$

(which follows from Theorem 3.4). For this reason the equation

$$(1+f_y^2)f_{xx} + (1+f_x^2)f_{yy} - 2f_xf_yf_{xy} = 0 \tag{4.36}$$

is referred to as the *equation of minimal surfaces* in $\mathbb{R}^3$.

A plane turns out to be the only regular minimal surface which is the graph of a function defined on the entire plane $xy$. This is a consequence of the fact that linear functions $f(x, y) = ax + by + c$ are the only regular solutions to equation (4.36) defined on the entire plane. This is *Bernstein's classical theorem*.

EXAMPLE 2. A CATENOID. If we write down the minimality condition for surfaces of revolution, then the resulting equation is solvable and its general solution has the form

$$\varphi(x) = a\cosh\Big(\frac{x}{a} + b\Big),$$

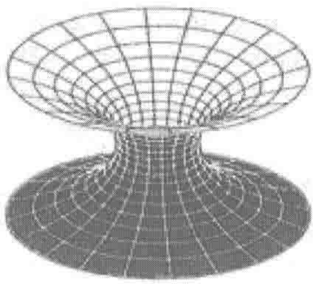

**Figure 4.5.** Catenoid.

with a nonzero constant $a$. These surfaces are obtained by rotation of the graph of the function $\varphi$ about the $x$-axis and are called *catenoids*. In the Weierstrass–Enneper representation they are specified by the functions $f = a/2$, $g = 1/z$ (or $\chi_1 = \sqrt{a}/\sqrt{2}$, $\chi_2 = \sqrt{a}/(\sqrt{2}z)$), where $z \in \mathbb{C} \setminus \{0\}$. The function $z$ is defined on the plane with a deleted point, but the integrals involved in the Weierstrass–Enneper formulas do not depend on the choice of the integration path. The explicit formulas are

$$x^1 = a\cosh u \cos v, \qquad x^2 = -a\cosh u \sin v, \qquad x^3 = au,$$

where $z = e^{u+iv}$.

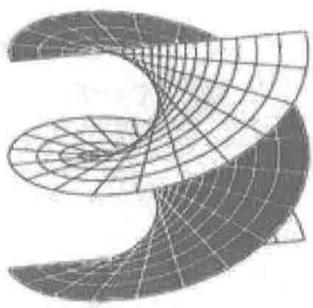

**Figure 4.6.** Helicoid.

EXAMPLE 3. A HELICOID. This surface is described by a straight line $l$ that rotates at a constant angular velocity around a fixed axis, intersects the axis at the right angle, with the intersection point uniformly moving along the axis.

The *helicoids* are specified by the functions $f = ia/2, g = 1/z$ (or $\chi_1 = \sqrt{ai}/\sqrt{2}$, $\chi_2 = \sqrt{ai}/(\sqrt{2}z)$), where $a \neq 0$ is a constant. Here $z \in \mathbb{C} \setminus \{0\}$, but now the integrals in equation (4.35) depend on the integration path. To achieve uniqueness, we remove from $\mathbb{C}$ the negative real half-line $\mathbb{R}_-$ given by the formulas $x \leq 0$, $y = 0$. Then we obtain a single turn of the helicoid corresponding to one rotation of $l$ through the angle $\pi$. The entire helicoid is obtained by analytic continuation: we must glue the upper boundary of

one copy of $\mathbb{C}\setminus\mathbb{R}_-$ to the lower boundary of the other copy, etc. The explicit formulas for the helicoid are

$$x^1 = a\sinh v \sin u, \qquad x^2 = -a\sinh v\cos u, \qquad x^3 = au,$$

where $z = e^{-i(u+iv)} = e^{v-iu}$. This implies that the helicoid is specified by the equation

$$\tan\frac{x^3}{a} = -\frac{x^1}{x^2}.$$

Note that the functions $\chi_1$ and $\chi_2$ for the catenoid and helicoid (with the same value of $a$) differ by the factor $\sqrt{i}$. In general, if minimal surfaces $r$ and $\tilde{r}$ are such that

$$\frac{\partial r}{\partial z} = e^{i\alpha}\frac{\partial \tilde{r}}{\partial z},$$

where $\alpha$ is a real constant, then they are said to be *associated*. Formula (4.33) implies that associated surfaces are locally isometric. For the catenoid and helicoid we have $\alpha = \pi/2$, and the local isometry is given by the mapping $(u, v) \to (v, u)$. However, globally they are different: the catenoid is an imbedded cylinder, whereas the helicoid is an imbedded plane. The latter statement follows from the fact that the helicoid may be specified by another pair of functions, $f = ie^{-z}$, $g = z$, which are defined on the entire complex plane.

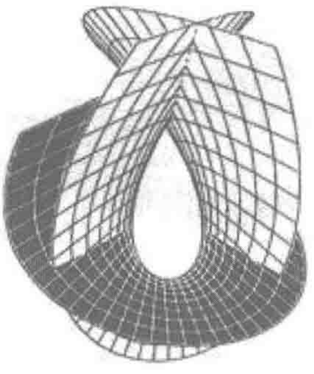

**Figure 4.7.** Enneper's surface.

EXAMPLE 4. ENNEPER'S SURFACE. Although this surface can be specified very simply: $f = 1$, $g = z$ (or $\chi_1 = 1$, $\chi_2 = z$), it is self-intersecting, and so it is immersed but not imbedded.

## Exercises to Chapter 4

1. Show that the only minimal surfaces of revolution are the plane and catenoids.

2. Let $dl^2 = du^2 + G(u)\,dv^2$ be the first fundamental form of a surface. Find its Gaussian curvature and construct explicitly a conformal parameter on this surface.

3. Show that the stereographic projection of a two-dimensional sphere onto the plane maps each planar section of the sphere (i.e., the circle in the intersection of the sphere with a plane) either into a circle or into a straight line.

4. Find the Riemannian metric on the one-sheeted hyperboloid induced by its imbedding into $\mathbb{R}^{1,2}$ as a pseudosphere of real radius.

5. Show that any surface (not necessarily minimal) may be represented by the Weierstrass–Enneper formulas (4.34). The functions $\chi_1, \chi_2$ must then satisfy the equation

$$\begin{pmatrix} U & \partial \\ -\bar{\partial} & U \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = 0,$$

where $\chi_1 = \sqrt{i}\psi_1, \chi_2 = \overline{\sqrt{i}\psi_2}$, the metric has the form

$$(|\psi_1|^2 + |\psi_2|^2)^2\, dz\, d\bar{z} = e^{2\alpha}\, dz\, d\bar{z},$$

and the real potential $U$ is expressed in terms of the metric and mean curvature $H$ by the formula

$$U = \frac{He^{\alpha}}{2}.$$

For $H = 0$ this representation turns into the Weierstrass–Enneper formulas for minimal surfaces.

*Chapter 5*

# Smooth Manifolds

## 5.1. Smooth manifolds

**5.1.1. Topological and metric spaces.** For completeness, we state here basic definitions and facts of the theory of topological and metric spaces.

A set $X$ is said to be equipped with a *topology* if there is a family of its subsets, which are called *open* sets, satisfying the following requirements:

1) the union of any family of open sets is open;

2) the intersection of finitely many open sets is open;

3) the set $X$ and its empty subset, i.e., the set containing no points of $X$, are open.

The set $X$ equipped with a topology is called a *topological space*. Any open set containing a point $x \in X$ is called a *neighborhood* of this point.

A set $V \subset X$ is said to be *closed* if its complement, i.e., the set $U = X \setminus V$, is open. It follows from the properties of open sets that:

1) the intersection of any family of closed sets is closed;

2) the union of finitely many closed sets is closed;

3) the space $X$ and its empty subset are closed.

Any subset $Y$ of a topological space $X$ is equipped in a natural way with the *induced* topology: a set $V \subset Y$ is said to be open if it is the intersection of $Y$ with an open set $U \subset X$, $V = U \cap Y$ (see Figure 5.1). The set $Y$ with this topology is called a *subspace* of the topological space $X$.

Moreover, for each subset $Y \subset X$ its *closure* $\bar{Y}$ is defined. It consists of points $x \in X$ such that any neighborhood of $x$ has a nonempty intersection with $Y$. Obviously, $Y \subset \bar{Y}$ and the set $\bar{Y}$ is closed.

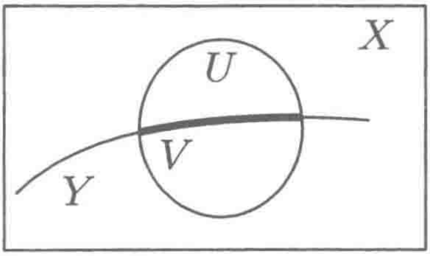


**Figure 5.1.** Induced topology.

Let $X$ and $Y$ be two topological spaces. A mapping $f\colon X \to Y$ is said to be *continuous* if the inverse image $f^{-1}(U)$ of any open set $U \subset Y$ is an open set in $X$. For mappings of Euclidean spaces $\mathbb{R}^n \to \mathbb{R}^m$ this definition coincides with the classical definition.

If a mapping of topological spaces $f\colon X \to Y$ is continuous, one-to-one, and its inverse $f^{-1}\colon Y \to X$ is also continuous, then $f$ is called a *homeomorphism* and the spaces $X$ and $Y$ are said to be *homeomorphic* or *topologically equivalent*. From the topological point of view such spaces are indistinguishable and have the same properties, unless they are endowed with other structures besides topology.

A set $X$ is a *metric space* if for any two points (elements of $X$) a distance between them, or a *metric*, is defined. This means that there is a nonnegative function

$$\rho\colon X \times X \to \mathbb{R}, \quad \rho \geq 0,$$

that associates with each pair of points $x, y \in X$ the *distance* $\rho(x, y)$ between them satisfying the following conditions:

1) $\rho(x, y) = \rho(y, x)$;

2) the *triangle inequality* holds:

$$\rho(x, z) \leq \rho(x, y) + \rho(y, z) \quad \text{for any } x, y, z \in X;$$

3) $\rho(x, y) = 0$ if and only if $x = y$.

The *open ball* $B(x, \varepsilon)$ of radius $\varepsilon$ with center at $x \in X$ is the set of all points of $X$ at distance less than $\varepsilon$ from $x$:

$$B(x, \varepsilon) = \{y \in X\colon \rho(x, y) < \varepsilon\}.$$

A subset $U \subset X$ is said to be open if together with any point $x_0 \in X$ it contains a sufficiently small open ball centered at $x_0$ (of radius depending on $x_0$). The open sets so defined specify a topology on the metric space $X$.

When we say that a topological space is a metric space, we mean that the topology is specified by a metric.

EXAMPLE. Euclidean spaces $\mathbb{R}^n$ are metric spaces with distance determined by the formula $\rho(x,y) = \sqrt{\sum_{i=1}^n (x^i - y^i)^2}$, where $x = (x^1, \dots, x^n)$ and $y = (y^1, \dots, y^n)$ are points of $\mathbb{R}^n$.

A metric $\rho_X$ on $X$ induces the metric $\rho_Y(x,y) = \rho_X(x,y)$ for $x, y \in Y$ on every subset $Y \subset X$, thus making it a metric space. The topology specified by the metric $\rho_Y$ is induced by the topology of $X$.

A mapping $f: X \to Y$ of metric spaces is called an *isometry* if it preserves the distances between points, i.e., $\rho_X(x,y) = \rho_Y(f(x), f(y))$, and there exists the inverse mapping $f^{-1}: Y \to X$. In this case the spaces $X$ and $Y$ are said to be *isometric* or *metrically equivalent*. Obviously, such spaces are homeomorphic.

The following definitions restate the definitions known from a course of calculus:

1) A sequence of points $\{x_n\}$ in a metric space is a *Cauchy sequence* if for any positive number $\varepsilon > 0$ there exists a positive integer $N = N(\varepsilon)$ such that $\rho(x_m, x_n) < \varepsilon$ for $m, n > N(\varepsilon)$.

2) A point $x_\infty$ is called the limit of a sequence $\{x_n\}$, $\lim_{n\to\infty} x_n = x_\infty$, if $\lim_{n\to\infty} \rho(x_\infty, x_n) = 0$.

If every Cauchy sequence in a metric space $X$ converges to a point in $X$, then the space is said to be *complete*.

A mapping of metric spaces $f: X \to Y$ is *continuous at a point* $x \in X$ if $\lim_{n\to\infty} x_n = x$ implies that $\lim_{n\to\infty} f(x_n) = f(x)$. A mapping $f$ of metric spaces is called continuous if it is continuous at every point.

A topological space $X$ is said to be *compact* if each covering $\{U_\alpha\}$ of it by open sets, $X = \bigcup_\alpha U_\alpha$, contains a finite subcovering, $X = \bigcup_{i=1}^N U_{\alpha_i}$.

A subset $Y \subset X$ of a topological space is said to be *compact* if it is compact as a topological space with induced topology.

The following lemma will enable us to establish compactness of many spaces.

**Lemma 5.1.** *Let $X$ be a compact topological space. Then:*

*1) If $f: X \to Y$ is a continuous mapping on the entire space $Y$, then the space $Y$ is compact.*

*2) If $Z \subset X$ is a closed subset of $X$, then it is compact as a topological space with induced topology.*

**Proof.** 1. Let $\{U_\alpha\}$ be a covering of $Y$ by open sets. Since the mapping $f$ is continuous, the sets $f^{-1}(U_\alpha)$ are open and cover the space $X$. Compactness of $X$ implies that this covering contains a finite subcovering $f^{-1}(U_j)$, $j = 1, \dots, k$. Now it remains to note that the sets $U_j$, $j = 1, \dots, k$, cover the

image of $f$, which is the space $f(X) = Y$. Therefore, any open covering of $Y$ contains a finite subcovering. Hence the image of a compact space under a compact mapping is also compact.

2. Let $\{U_\alpha\}$ be a covering of $Z$ by sets that are open in the induced topology. This means that they have the form $U_\alpha = Z \cap V_\alpha$, where $V_\alpha$ is an open set in $X$. The complement of $Z$ is the open set $X \setminus Z$. The sets $\{V_\alpha\}$ together with this complement form an open covering of $X$. Now we select a finite subcovering $X \setminus Z, V_1, \dots, V_k$ and observe that the sets $U_1 = Z \cap V_1, \dots, U_k = Z \cap V_k$ form a finite covering of $Z$. Hence each open covering of $Z$ contains a finite subcovering, which completes the proof. □

**Corollary 5.1.** *Let $X$ be a compact topological space and $f\colon X \to \mathbb{R}$ a continuous function. Then for any $c \in \mathbb{R}$ the level set $f^{-1}(c) = \{x \in X\colon f(x) = c\}$ is compact.*

It is well known from a course of calculus that a subset $X \subset \mathbb{R}^n$ is compact if and only if $X$ is closed and bounded in $\mathbb{R}^n$.

Another fact from calculus holds for all compact spaces.

**Theorem 5.1.** *Every continuous function on a compact space attains its minimum and maximum.*

A continuous mapping $f\colon X \to Y$ is said to be *proper* if for any compact set $K \subset Y$ the inverse image $f^{-1}(K)$ is also compact.

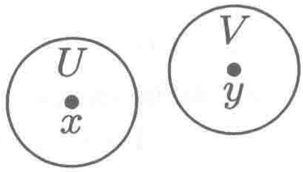


**Figure 5.2.** Disjoint neighborhoods.

If any two different points $x$ and $y$ of a topological space $X$ possess disjoint neighborhoods $U$ and $V$,

$$U \cap V = \varnothing, \quad x \in U, \quad y \in V,$$

then $X$ is called a *Hausdorff space* (see Figure 5.2).

**Theorem 5.2.** *Every metric space is a Hausdorff space.*

Indeed, if $\rho(x, y) = r > 0$, then the open balls $B(x, r/3)$ and $B(y, r/3)$ with centers at these points are disjoint.

The existence of a metric is the most effective way to prove that a topological space is Hausdorff. In what follows we will consider only such spaces, usually without special mention.

A topological space $X$ is *connected* if it cannot be represented as a union of two nonempty disjoint subsets that are open and closed at the same time.

A topological space $X$ is *arcwise connected* if any two of its points $x_1, x_2 \in X$ can be joined by a continuous line, i.e., there exists a continuous mapping $f\colon [0,1] \to X$ such that $f(0) = x_1$ and $f(1) = x_2$.

A connected space $X$ is *simply connected* if each continuous mapping $f\colon S^1 \to X$ of the unit circle $\{|z| = 1\} \subset \mathbb{C}$ into $X$ may be extended to a continuous mapping of the unit disk $\{|z| \leq 1\}$ into $X$.

Another way to define simple connectedness uses the notion of homotopy.

Continuous mappings $f_0\colon X \to Y$ and $f_1\colon X \to Y$ are *homotopic* if there exists a continuous mapping $F\colon X \times [0,1] \to Y$ that coincides with $f_0$ for $t = 0$ and with $f_1$ for $t = 1$, i.e., $f_0(x) = F(x,0)$, $f_1(x) = F(x,1)$. This mapping is called a *homotopy* between the mappings $f_0$ and $f_1$.

Similarly, two paths $r_0\colon [0,1] \to X$ and $r_1\colon [0,1] \to X$ from $x_0$ to $x_1$ are said to be *homotopic* if there exists a homotopy between these paths $F\colon [0,1] \times [0,1] \to X$ (i.e., $F(t,0) = r_0(t)$ and $F(t,1) = r_1(t)$) such that for any fixed $s \in [0,1]$ the mapping $F(\cdot, s)\colon [0,1] \to X$ specifies a path from $x_0 = F(0,s)$ to $x_1 = F(1,s)$.

It is easily seen that if we divide the circle $S^1 = \{|z| = 1\}$ into the upper and lower arcs specifying the paths from $f(1)$ to $f(e^{i\pi})$, then the homotopy between these paths specifies an extension of the mapping $f\colon S^1 \to X$ onto the entire disk $\{|z| \leq 1\}$. Therefore, we have the following lemma.

**Lemma 5.2.** *A space $X$ is simply connected if and only if for any two of its points $x_0$ and $x_1$, all paths from $x_0$ to $x_1$ are homotopic to each other.*

**5.1.2. On the notion of smooth manifold.** Topological spaces that have the structure of a Euclidean space in a neighborhood of each point are called manifolds. Now we will give a formal definition.

Let $M$ be a set of points. We say that $M$ carries a smooth *atlas* if there is a finite or countable family $\{U_\alpha\}$ of subsets of $M$ with the following properties:

1) These sets form a *covering*, i.e., $M$ lies in their union, $M \subset \bigcup_\alpha U_\alpha$.

2) The points of each set $U_\alpha$ are in a one-to-one correspondence, $U_\alpha \leftrightarrow V_\alpha$, with the points of a domain $V_\alpha \subset \mathbb{R}^n$. Therefore, one can introduce *local coordinates* $(x^1_\alpha, \dots, x^n_\alpha)$ in $U_\alpha$ by assigning to a point $x \in U_\alpha$ the coordinates of the corresponding point in $V_\alpha$.

3) In the intersection $U_\alpha \cap U_\beta$, the local coordinates $(x^1_\alpha, \dots, x^n_\alpha)$ and $(x^1_\beta, \dots, x^n_\beta)$ are related to each other by mutually inverse smooth changes of coordinates,

$$x^i_\alpha = x^i_\alpha(x^1_\beta, \dots, x^n_\beta), \qquad x^j_\beta = x^j_\beta(x^1_\alpha, \dots, x^n_\alpha), \quad i, j = 1, \dots, n, \tag{5.1}$$

with nonzero Jacobians,

$$\det\left(\frac{\partial x^i_\alpha}{\partial x^j_\beta}\right) \neq 0, \qquad \det\left(\frac{\partial x^i_\beta}{\partial x^j_\alpha}\right) \neq 0. \tag{5.2}$$

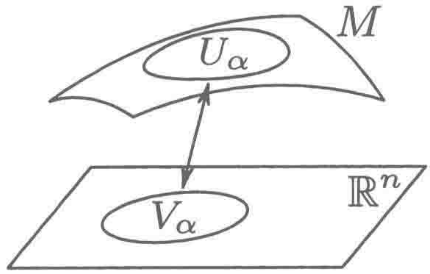


**Figure 5.3.** A chart.

The sets $U_\alpha$ that constitute a smooth atlas are called *charts*. An atlas specifies a topology: a set $U \subset M$ is open if the coordinates of the points in its intersection $U \cap U_\alpha$ with any chart $U_\alpha$ of the atlas form an open set in $\mathbb{R}^n$.

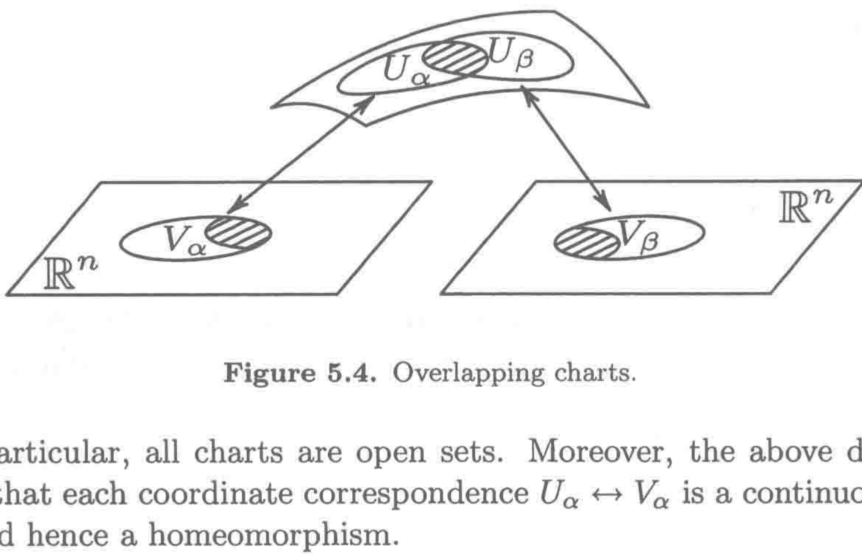


**Figure 5.4.** Overlapping charts.

In particular, all charts are open sets. Moreover, the above definition implies that each coordinate correspondence $U_\alpha \leftrightarrow V_\alpha$ is a continuous mapping, and hence a homeomorphism.

If the topology on $M$ is specified by a smooth atlas and one of the following two conditions holds:

1) $M$ is a metric space,

2) $M$ is a Hausdorff space,

then $M$ is called a *smooth manifold*.

The number $n$ is called the *dimension* of the manifold $M$, $\dim M = n$.

It can be shown that the two conditions in the definition of a smooth manifold are equivalent. Note that we specify an atlas first, and then specify the topology by means of this atlas.

For smooth manifolds the notions of connectedness and arcwise connectedness are equivalent.

As is shown by the following example, the condition that the space is Hausdorff is essential.

EXAMPLE. STRAIGHT LINE WITH A DOUBLE POINT. Consider two copies of a straight line, $\mathbb{R}_+$ and $\mathbb{R}_-$, with coordinates $x_+$ and $x_-$. Let us glue them along the complements of the points $x_+ = 0$ and $x_- = 0$, identifying the points with coordinates $x_+ = y$ and $x_- = y$ for any $y \in \mathbb{R}$, $y \neq 0$. The resulting set $X$ is covered by two charts, $U_+$ and $U_-$, with coordinates $x_+$ and $x_-$. The chart $U_+$ covers the whole set $X$, except for the point which was initially on the line $\mathbb{R}_-$ and had the coordinate $x_- = 0$. The chart $U_-$ is defined similarly, changing the signs $\pm \leftrightarrow \mp$ everywhere. We see that any two neighborhoods of the points $x_+ = 0$ and $x_- = 0$ intersect. Hence $X$ is not a Hausdorff space.

Now we will give the simplest examples of smooth manifolds.

EXAMPLES. 1. EUCLIDEAN SPACES $\mathbb{R}^n$ AND THEIR DOMAINS. When speaking about a Euclidean space or of a domain of this space, we mean that it is covered by a single chart with Euclidean coordinates.

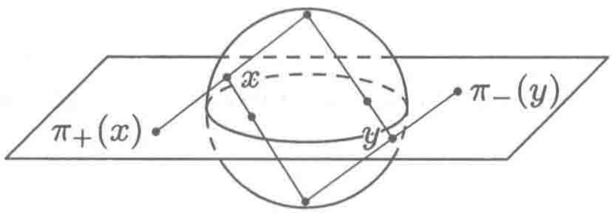


**Figure 5.5.** Stereographic projections.

2. SPHERES $S^n$. Consider the unit sphere $S^n$ in $\mathbb{R}^{n+1}$; it consists of all the points at unit distance from the origin. Take two opposite points on it: the "North Pole" $P_+ = (0, \ldots, 0, 1)$ and the "South Pole" $P_- = (0, \ldots, 0, -1)$. On the sphere $S^n \setminus P_+$ with deleted "North Pole", the stereographic projection $\pi_+$ is defined, which maps each point $x$ into the intersection point of the plane $x^{n+1} = 0$ with the straight line passing through the points $P_+$ and $x$,

$$\pi_+(x^1, \ldots, x^{n+1}) = (y_+^1, \ldots, y_+^n, 0), \quad (y_+^1, \ldots, y_+^n) = \frac{1}{1 - x^{n+1}} (x^1, \ldots, x^n).$$

In a similar way the stereographic projection $\pi_-$ of the sphere with deleted "South Pole" is defined:

$$\pi_-(x^1, \ldots, x^{n+1}) = (y_-^1, \ldots, y_-^n, 0), \quad (y_-^1, \ldots, y_-^n) = \frac{1}{1 + x^{n+1}} (x^1, \ldots, x^n).$$

Thus we obtain an atlas consisting of two charts, $U_+ = S^n \setminus P_+$ with coordinates $(y_+)$ and $U_- = S^n \setminus P_-$ with coordinates $(y_-)$. In the intersection of these charts the coordinates are related by smooth transition formulas: for any point $x$ they are inverse proportional and

$$|y_+||y_-| = 1.$$

The smooth manifold $S^n$ with such an atlas is called the $n$-dimensional *sphere.*

3. PROJECTIVE SPACES $\mathbb{R}P^n$. Denote by $\mathbb{R}P^n$ the set of all straight lines passing through the origin in $\mathbb{R}^{n+1}$. Each line intersects the unit sphere in two points, which can be interchanged by reflection $\sigma\colon x \to -x$. Let us construct an atlas on $\mathbb{R}P^n$. To this end, take the atlas $\{U_+, U_-\}$ on the sphere and construct a collection of open subsets of $U_\pm$ covering $S^n$ such that none of them overlaps with its image under the reflection $\sigma$. This can be done, for example, as follows:

$$U_{\pm,i,+} = U_\pm \cap \{x^i > 0\}, \qquad U_{\pm,i,-} = U_\pm \cap \{x^i < 0\}, \quad i = 1, \dots, n+1.$$

In these charts we retain the same coordinates $y_\pm$. Thus we obtain charts on $\mathbb{R}P^n$: each subset $U$ parametrizes the straight lines crossing the sphere at the points of $U$. These charts cover the entire space $\mathbb{R}P^n$, and there exist smooth changes of coordinates on their intersections. Hence we obtain a smooth atlas on the manifold $\mathbb{R}P^n$, which is called the $n$-dimensional *real projective space.*

4. PRODUCTS OF MANIFOLDS. TORI. Suppose we have two manifolds, $M^n$ with atlas $\{U_\alpha\}$ and $N^k$ with atlas $\{V_\beta\}$. Consider their direct product $M^n \times N^k$, which is the set of all pairs $(x, y)$, where $x \in M^n$ and $y \in N^k$. On this set we introduce the atlas consisting of all products of charts $\{U_\alpha \times V_\beta\}$ with local coordinates $(x^1_\alpha, \dots, x^n_\alpha, y^1_\beta, \dots, y^k_\beta)$. Thus we obtain a smooth manifold $M^n \times N^k$ of dimension $(n+k)$ called the *product of the manifolds* $M^n$ and $N^k$. This procedure may be applied repeatedly. In this way one obtains, e.g., the $n$-dimensional *torus* $T^n = S^1 \times \cdots \times S^1$, which is the product of $n$ copies of the circle $S^1$.

5. CONNECTED SUMS OF MANIFOLDS. Let $M_1$ and $M_2$ be smooth manifolds of the same dimension $n$. In each of them take a point, $x_1 \in M_1$ and $x_2 \in M_2$, and a coordinate neighborhood of this point. We may assume that $x_1 = 0$ and $x_2 = 0$ and that the neighborhoods contain balls $|x_j| \le 2r$, $j = 1, 2$. Now we delete the open ball $|x_j| < r$ from the corresponding manifold and glue the boundaries of the remaining manifolds along the boundary spheres $|x_1| = r$ and $|x_2| = r$, identifying the points with the same coordinates $x_1$ and $x_2$. The boundary spheres have neighborhoods whose closures are diffeomorphic to cylinders, say, $S^{n-1} \times [-1, 0]$ and $S^{n-1} \times [0, 1]$. Gluing

the boundaries joins them into the cylinder $S^{n-1} \times [-1, 1]$, to which smooth coordinates from the initial neighborhoods are extended (see Figure 5.6).

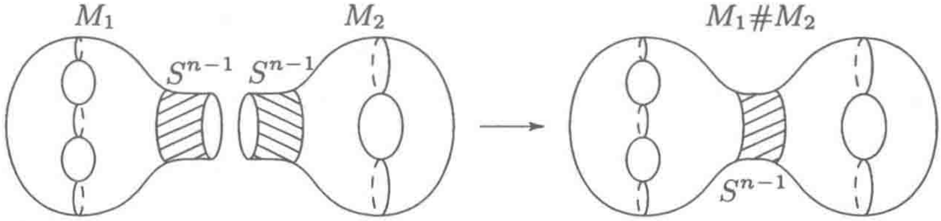


**Figure 5.6.** Connected sum of manifolds.

Thus we obtain a new manifold, which is denoted by

$$M_1 \# M_2$$

and is called the *connected sum* of the manifolds $M_1$ and $M_2$.

6. TANGENT BUNDLES. Let $M^n$ be a smooth manifold. Then the pairs $(x, v)$, where $x \in M^n$ and $v$ is a tangent vector at the point $x$, form a smooth manifold $TM^n$ of dimension $2n$. Indeed, to each chart $U_\alpha$ with coordinates $x^1_\alpha, \dots, x^n_\alpha$ on the manifold $M^n$ corresponds the chart $\tilde{U}_\alpha$ with coordinates $x^1_\alpha, \dots, x^n_\alpha, \xi^1_\alpha, \dots, \xi^n_\alpha$, where $\xi = \dot{x}$ is the tangent vector at $x$ to some curve (recall that all tangent vectors to $M^n$ are specified in this way). The transition functions between the charts have the form

$$x^i_\beta = x^i_\beta(x_\alpha), \qquad \xi^j_\beta = \frac{\partial x^j_\beta}{\partial x^i_\alpha}(x_\alpha)\,\xi^i_\alpha.$$

The manifold $TM^n$ is called the *tangent bundle* to $M^n$.

**5.1.3. Smooth mappings and tangent spaces.** Having introduced a smooth atlas, we can define the notion of a smooth function on a manifold.

A function $f\colon M^n \to \mathbb{R}$ on a manifold is *smooth* if in a neighborhood of each point of the smooth manifold $M^n$ it is a smooth function of local coordinates, $f = f(x^1_\alpha, \dots, x^n_\alpha)$. Since there are smooth invertible changes of coordinates (5.1) in the intersections of charts, a smooth function relative to some coordinates is smooth relative to others. Furthermore, by the chain rule,

$$\frac{\partial f}{\partial x^i_\beta} = \frac{\partial f}{\partial x^j_\alpha}\frac{\partial x^j_\alpha}{\partial x^i_\beta}. \tag{5.3}$$

Hence a smooth function is well defined.

Obviously, continuous functions on manifolds can also be defined by the condition that, written in local coordinates, they are continuous functions of these coordinates.

If $M^n$ and $N^k$ are smooth manifolds, then a mapping $F\colon M^n \to N^k$ is said to be *smooth* if, written in local coordinates on $M^n$ and $N^k$, it is a smooth mapping of a domain in $\mathbb{R}^n$ into a domain in $\mathbb{R}^k$,

$$y^1 = F^1(x^1, \dots, x^n), \ \dots, \ y^k = F^k(x^1, \dots, x^n).$$

Two smooth manifolds $M^n$ and $N^k$ are said to be *diffeomorphic* if there exist mutually inverse smooth mappings $f\colon M^n \to N^k$ and $f^{-1}\colon N^k \to M^n$ (in which case, obviously, $n = k$). Such mappings are called *diffeomorphisms*. We will not distinguish between diffeomorphic manifolds.

Suppose there are two different smooth atlases on the same topological space $M^n$. These atlases are *equivalent* or specify the same *smooth structure* if the sets of smooth functions relative to each of the atlases coincide. This means that the identity mappings $M_1^n \to M_2^n$ and $M_2^n \to M_1^n$ are smooth. Here the subscript indicates the number of the atlas specifying the smooth structure on $M^n$. The manifolds with equivalent atlases are diffeomorphic as smooth manifolds.

Let $r(t)$ be a smooth curve on a manifold $M^n$. In local coordinates $x_\alpha^i$ it is specified by smooth functions,

$$r(t) = (x_\alpha^1(t), \dots, x_\alpha^n(t)).$$

Suppose that it passes through a point $x \in M^n$ at time $t = 0$. Its velocity vector $\xi = \dot{r}(0)$ at this point is called a *tangent vector* to the manifold $M^n$ at the point $x$. In the coordinates $x_\alpha^i$ it is specified by an $n$-tuple of the form

$$\xi = \Big(\frac{dx_\alpha^1(0)}{dt}, \dots, \frac{dx_\alpha^n(0)}{dt}\Big).$$

In another coordinate system $x_\beta^j$ the same vector is specified by another $n$-tuple,

$$\tilde{\xi} = \Big(\frac{dx_\beta^1(0)}{dt}, \dots, \frac{dx_\beta^n(0)}{dt}\Big).$$

These representations are related by the transition formulas

$$\tilde{\xi}^i = \frac{\partial x_\beta^i}{\partial x_\alpha^j}\, \xi^j, \tag{5.4}$$

which, like (5.3), are obtained by the chain rule. By conditions (5.2), the transformation $\xi \to \tilde{\xi}$ is invertible. We consider two tangent vectors $\xi$ and $\eta$ at the point $x$ to be equal if they are specified by the same $n$-tuple in some coordinate system. The above arguments show that this notion is well defined.

All the tangent vectors at the point $x$ form a vector space with the usual operations of componentwise addition,

$$\xi + \eta = (\xi^1 + \eta^1, \dots, \xi^n + \eta^n),$$

and multiplication by a real number,

$$\lambda\xi = (\lambda\xi^1, \dots, \lambda\xi^n).$$

This space is called the *tangent space* to the manifold at the point $x$ and is denoted by $T_x M^n$. Obviously, its dimension is equal to the dimension of the manifold.

In order to specify a scalar product

$$\langle \xi, \eta \rangle = g_{ij}\xi^i\eta^j$$

on the tangent space, we must assign the coefficients $g_{ij}^{\alpha}$ (which are functions of coordinates) so that the value of the scalar product be invariant with respect to coordinate changes,

$$g_{kl}\xi^k\eta^l = \tilde{g}_{ij}\tilde{\xi}^i\tilde{\eta}^j = \tilde{g}_{ij}\Big(\frac{\partial x_\beta^i}{\partial x_\alpha^k}\,\xi^k\Big)\Big(\frac{\partial x_\beta^j}{\partial x_\alpha^l}\,\eta^l\Big) = \Big(\tilde{g}_{ij}\,\frac{\partial x_\beta^i}{\partial x_\alpha^k}\,\frac{\partial x_\beta^j}{\partial x_\alpha^l}\Big)\xi^k\eta^l.$$

To satisfy this requirement, $g_{ij}$ must transform under the coordinate changes by the rule

$$g_{kl} = \tilde{g}_{ij}\,\frac{\partial x_\beta^i}{\partial x_\alpha^k}\,\frac{\partial x_\beta^j}{\partial x_\alpha^l}.$$

If at every point of a manifold $M^n$ there is a positive definite scalar product, then the manifold is said to be equipped with a *Riemannian metric* $g_{ij}$. Such a manifold is called *Riemannian.*

By means of the Riemannian metric we assign to each smooth curve $r(t)$, $a \le t \le b$, its length

$$l(r) = \int_a^b \sqrt{\langle \dot{r}, \dot{r} \rangle}\,dt = \int_a^b |\dot{r}|\,dt.$$

Let the Riemannian manifold $M^n$ be connected. The Riemannian metric determines a metric on $M^n$ in the usual sense, i.e., as the distance between points,

$$\rho(x, y) = \inf_{\substack{r\colon [a,b]\to M^n,\\ r(a)=x,\, r(b)=y}} l(r),$$

where the infimum is taken over all smooth curves joining $x$ and $y$. The space $M^n$ with such a distance $\rho$ is a metric space.

Every smooth mapping of manifolds $F\colon M^n \to N^k$ determines a linear mapping of tangent spaces called the *differential* $F_*$. Namely, if $r(t)$ is a smooth curve on $M^n$ and $\dot{r}(t)$ is its tangent vector at a point $r(t)$, then

the mapping takes $\dot{r}(t)$ into the tangent vector $\dot{s}(t)$ at the point $s(t)$, where $s(t) = F(r(t))$ is a smooth curve in the coordinates of $N^k$.

If $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^k)$ are local coordinates on $M^n$ and $N^k$, respectively, then we have

$$r(t) = (x^1(t), \ldots, x^n(t)), \quad \dot{r} = (\dot{x}^1(t), \ldots, \dot{x}^n(t)),$$

and setting

$$F(x^1, \ldots, x^n) = (y^1(x^1, \ldots, x^n), \ldots, y^k(x^1, \ldots, x^n)),$$

we obtain

$$s(t) = (y^1(x^1(t), \ldots, x^n(t)), \ldots, y^k(x^1(t), \ldots, x^n(t))).$$

Therefore, the differential $F_*$, which assigns the vector $\dot{s}$ to the vector $\dot{r}$, has the form

$$(\dot{x}^1(t), \ldots, \dot{x}^n(t)) \to \Big(\frac{\partial y^1}{\partial x^j}\, \dot{x}^j(t), \ldots, \frac{\partial y^k}{\partial x^j}\, \dot{x}^j(t)\Big),$$

and so it is specified by the Jacobi matrix

$$J = \Big(\frac{\partial y^i}{\partial x^j}\Big).$$

Let $F\colon M^n \to N^k$ be a smooth mapping of manifolds. If at every point $x \in M^n$ the differential $F_*$ is an imbedding of the tangent space $T_x M^n$ into the space $T_{F(x)} N^k$, then the mapping $F$ is called an *immersion*.

If an immersion $F\colon M^n \to N^k$

a) maps different points of $M^n$ into different points of $N^k$ (i.e., $F(x) \neq F(y)$ for $x \neq y$) and

b) is a proper mapping (i.e., the image of any compact set is also compact),

then $F$ is called an *imbedding* and its image $F(M^p)$ is called a *submanifold* of the manifold $N^q$.

EXAMPLE. Let $T^n = S^1 \times \cdots \times S^1$ be the $n$-dimensional torus. On each factor there is an angular coordinate $\varphi_i$ defined modulo $2\pi$. Consider a mapping $F\colon \mathbb{R} \to T^n$ of the form $F(t) = (\alpha_1 t, \ldots, \alpha_n t)$. It is an immersion and takes different points into different points, $F(x) \neq F(y)$ for $x \neq y$. Suppose that the numbers $\alpha_i$, $i = 1, \ldots, n$, are linearly independent over the field of rational numbers $\mathbb{Q}$ (for example, for $n = 2$ this means that the ratio $\alpha_1/\alpha_2$ is irrational). Then the "winding" $F(\mathbb{R})$ is everywhere dense on the torus, i.e., the closure of the image of the mapping $F$ coincides with the entire torus, $\overline{F(\mathbb{R})} = T^n$. This mapping is not proper, although it has all the other properties of an imbedding.

If a mapping $M^n$ is compact, then any sequence of points $\{x_i\}$ in $M^n$ has an *accumulation point.* This means that it contains a subsequence $\{x_{i_k}\}$ converging in some chart to a point $x_\infty \in M^n$, which is called an accumulation point. This implies that if $M^n$ is a compact manifold, then any imbedding of it is proper.

In each tangent space one can introduce an orientation. If the orientations are compatible, then the manifold is said to be oriented. Let us give a formal definition: we say that a manifold $M^n$ is *oriented* (or that an *orientation* is given on $M^n$) if $M^n$ is covered by an atlas such that all the Jacobians of the transition functions are positive,

$$\det\left(\frac{\partial x_\alpha^i}{\partial x_\beta^j}\right) > 0.$$

Then, by definition, the basis of tangent vectors written in these coordinates as $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ is positively oriented at each point.

If a manifold can be oriented, it is said to be *orientable.*

**5.1.4. Multidimensional surfaces in $\mathbb{R}^n$. Manifolds with boundary.** As a particular case of smooth manifolds, we will consider regular multidimensional surfaces in Euclidean spaces.

A set $M^k \subset \mathbb{R}^n$ is a *k-dimensional surface* (or a $k$-dimensional submanifold) in $\mathbb{R}^n$ if for any point $x \in M^k$ there is a neighborhood $U$ of this point such that in this neighborhood, $M^k$ is specified as the set of zeros of a smooth mapping $F\colon U \to \mathbb{R}^{n-k}$ (a collection of $n-k$ functions),

$$(x^1, \ldots, x^n) \xrightarrow{F} (F_1(x^1, \ldots, x^n), \ldots, F_{n-k}(x^1, \ldots, x^n)).$$

A surface is said to be *nonsingular* or *regular* if all its points are nonsingular, i.e., the rank of the matrix $\left(\frac{\partial F_i}{\partial x^j}\right)$ at any point is $(n-k)$.

Let $x \in M^k$. We may assume without loss of generality that in a neighborhood of this point the minor

$$\det\left(\frac{\partial F_i}{\partial x^j}\right), \quad k+1 \le j \le n,$$

does not vanish. By the implicit function theorem there are smooth functions $f_1, \ldots, f_{n-k}$ defined in a neighborhood of $x$ such that

$$F_1(x) = \cdots = F_{n-k}(x) = 0$$

if and only if

$$\begin{aligned} x^{k+1} &= f_1(x^1, \ldots, x^k), \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ x^n &= f_{n-k}(x^1, \ldots, x^k). \end{aligned} \tag{5.5}$$

Therefore, in a small neighborhood of $x$ the surface $M^k$ is the graph of the smooth mapping

$$(x^1, \dots, x^k) \to (f_1, \dots, f_{n-k}).$$

Let us introduce new variables in a neighborhood of $x$,

$$z^1 = x^1, \dots, z^k = x^k,$$
$$z^{k+1} = x^{k+1} - f_1(x^1, \dots, x^k), \dots, z^n = x^n - f_{n-k}(x^1, \dots, x^k).$$

The determinant of the matrix $\left(\frac{\partial z^i}{\partial x^j}\right)$ is identically equal to one; hence by the inverse function theorem, $z^1, \dots, z^n$ are coordinates in a neighborhood of the point $x$ in $\mathbb{R}^n$. Then the $k$-dimensional surface $M^k$ is specified in this neighborhood by the equations

$$z^{k+1} = \dots = z^n = 0.$$

In this case $z^1, \dots, z^k$ are *local coordinates* on the surface (in a neighborhood of a given point).

Let $r(t) = (x^1(t), \dots, x^n(t))$ be a smooth curve on the surface $M^k$. Then relations (5.5) imply that its velocity vector $\dot{r}(t)$ at a point of the surface has the form

$$\dot{r}(t) = \dot{x}^1 e_1 + \dots + \dot{x}^k e_k,$$

where

$$e_1 = \left(1, 0, \dots, 0, \frac{\partial f_1}{\partial x^1}, \dots, \frac{\partial f_{n-k}}{\partial x^1}\right),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$e_k = \left(0, 0, \dots, 1, \frac{\partial f_1}{\partial x^k}, \dots, \frac{\partial f_{n-k}}{\partial x^k}\right).$$

The vectors $e_1, \dots, e_k$ are linearly independent and form a basis in the $k$-dimensional *tangent plane* to the surface at this point. In the basis $e_1, \dots, e_k$, the velocity vector has coordinates $(\dot{x}^1, \dots, \dot{x}^k)$.

If

$$F_1(x^1, \dots, x^n) = \dots = F_{n-k}(x^1, \dots, x^n) = 0$$

are equations specifying the surface, then the coordinates $\xi = (\xi^1, \dots, \xi^n)$ of any tangent vector at a point $x \in M^k$ satisfy the equations

$$\frac{\partial F_i(x)}{\partial x^j} \xi^j = 0, \quad i = 1, \dots, n-k. \tag{5.6}$$

It is easily seen that any vector satisfying these equations is a tangent vector to the surface at the point $x$.

The following statement is obvious.

**Lemma 5.3.** *A point $x$ is a nonsingular point of the surface $F_1(x) = \cdots = F_{n-k}(x) = 0$ if and only if equation (5.6) for the tangent vectors $\xi = (\xi^1, \dots, \xi^n)$ specifies a $k$-dimensional vector subspace. If a point $x$ is singular, then these equations specify a subspace of dimension greater than $k$.*

Indeed, the dimension of the space of solutions to the linear system (5.6) is $n - \operatorname{rank}\left(\frac{\partial F_i(x)}{\partial x^j}\right)$, and the matrix $\left(\frac{\partial F_i(x)}{\partial x^j}\right)$ has maximal rank ($= n-k$) at nonsingular points and only them.

Thus we have proved the following theorem.

**Theorem 5.3.** *For any regular smooth $k$-dimensional surface $M^k$ in $\mathbb{R}^n$ there is a smooth atlas such that all coordinate functions $x^1, \dots, x^n$ are smooth functions on the surface.*

This smooth structure on the surface is canonical: all atlases for which the coordinate functions are smooth yield diffeomorphic manifolds. Indeed, let $M_1^k$ and $M_2^k$ be the same regular surface with two different atlases. If for both atlases the coordinate functions are smooth, then the identity mappings $M_1^k \to M_2^k$ and $M_2^k \to M_1^k$ are diffeomorphisms. In what follows we will always assume that a regular surface is equipped with this smooth structure.

Examples. 1. Unit spheres $S^n$ are regular surfaces in $\mathbb{R}^{n+1}$ specified by the equations

$$(x^1)^2 + \cdots + (x^{n+1})^2 = 1.$$

2. Consider regular surfaces in $\mathbb{R}^{10} = \mathbb{C}^5$ specified by the equations

$$z_1^{6k-1} + z_2^3 + z_3^2 + z_4^2 + z_5^2 = 0, \qquad |z_1|^2 + |z_2|^2 + |z_3|^2 + |z_4|^2 + |z_5|^2 = 1$$

for $k = 1, \dots, 28$. Here $z_1, \dots, z_5$ are complex linear coordinates in $\mathbb{C}^5$. The first equation is complex-valued and consists of two real-valued equations (for real and imaginary parts of the polynomial). It can be shown that all these regular surfaces are pairwise nondiffeomorphic, but all of them are homeomorphic to the (7-dimensional) sphere $S^7$. Furthermore, any smooth manifold homeomorphic to $S^7$ is diffeomorphic to one of these surfaces. The proofs of these results (due to Milnor) require deep methods of classical topology.

In connection with Example 2, we will state another surprising fact. If a manifold $M^n$ is homeomorphic to the Euclidean space $\mathbb{R}^n$ for $n \neq 4$, then it is diffeomorphic to this space. This is also a result of classical topology. Using the methods of the calculus of variations on manifolds discovered in theoretical physics (the so-called instanton phenomenon due to Polyakov–Belavin–Schwarz–Tyupkin), it was proved that this result does not hold for $n = 4$ (Donaldson): there are infinitely many pairwise nondiffeomorphic manifolds homeomorphic to $\mathbb{R}^4$.

**Theorem 5.4.** *Let $M^k \subset \mathbb{R}^n$ be a $k$-dimensional regular surface specified everywhere by the equations*

$$F_1(x) = \cdots = F_{n-k}(x) = 0.$$

*Then it is orientable.*

**Proof.** At any point $x \in M^k$, the vectors

$$\operatorname{grad} F_1 = \Big(\frac{\partial F_1}{\partial x^1}, \dots, \frac{\partial F_1}{\partial x^n}\Big), \ \dots, \ \operatorname{grad} F_{n-k} = \Big(\frac{\partial F_{n-k}}{\partial x^1}, \dots, \frac{\partial F_{n-k}}{\partial x^n}\Big)$$

are linearly independent and orthogonal to the tangent plane to $M^k$. We can specify orientation in the tangent planes to $M^k$ as follows: a basis $e_1, \dots, e_k$ is positively oriented if the vectors $e_1, \dots, e_k, \operatorname{grad} F_1, \dots, \operatorname{grad} F_{n-k}$ form a positively oriented basis in the Euclidean space $\mathbb{R}^n$. This proves the theorem. □

**Corollary 5.2.** *The spheres $S^n$ are orientable for all $n \geq 1$.*

Similarly to surfaces in $\mathbb{R}^n$, one can define surfaces in arbitrary smooth manifolds. For example, let $f: M^n \to \mathbb{R}$ be a smooth function and suppose that its gradient $(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n})$ does not vanish anywhere on the set of its zeros. Since the gradients of a function in different coordinate systems are related by formula (5.3), this condition holds regardless of the choice of coordinates. Denote by $Q^{n-1}$ the set of zeros of $f$. In a neighborhood of each point of $Q^{n-1}$ the manifold $M^n$ has the structure of a Euclidean space; hence we can choose a smooth atlas on $Q^{n-1}$ such that all the coordinate functions $x^1, \dots, x^n$ are smooth functions on $Q^{n-1}$.

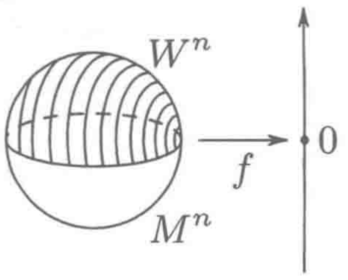


**Figure 5.7.** Manifold with a boundary.

The closed set $W^n \subset M^n$ specified by the inequality $f(x) \geq 0$ is called a smooth *manifold with a boundary* $Q^{n-1} = \partial W^n$.

In a neighborhood of any boundary point $x$, it is similar to the half-space $x^k \geq 0$ with the point $x$ lying on the hyperplane $x^k = 0$.

If a manifold $M^n$ is compact and has no boundary, it is called *closed*.

**5.1.5. Partition of unity. Manifolds as multidimensional surfaces in Euclidean spaces.** For some purposes it is convenient to use atlases of a special form. Namely, one says that on a compact manifold $M^n$ there is an atlas "with thick boundary" consisting of charts $U_\alpha$, $\alpha = 1, \dots, k$, if

1) in each chart the local coordinates $x^i_\alpha, i = 1, \dots, n$, run over the unit ball $\sum_i (x^i_\alpha)^2 < 1$;

2) the balls $W_\alpha \subset U_\alpha$ of radius $1/2$, $\sum_i (x^i_\alpha)^2 < 1/2$, also form a covering of $M^n$, i.e., the charts $U_\alpha, \alpha = 1, \dots, k$, cover the manifold $M^n$ "with excess".

Obviously, an atlas of charts "with thick boundary" can be chosen for any compact manifold.

We say that we have a *partition of unity* on $M^n$ if there is a family of smooth functions $\varphi_\alpha$ and a covering of the manifold by coordinate domains $U_\alpha$ such that

1) each function $\varphi_\alpha$ is concentrated in $U_\alpha$, i.e., $\varphi_\alpha(x) = 0$ for $x \in M^n \backslash U_\alpha$;

2) $0 \le \varphi_\alpha \le 1$ for all $\alpha$;

3) at any point $x$ there are only finitely many nonzero functions $\varphi_\alpha$ and

$$\sum_\alpha \varphi_\alpha(x) \equiv 1.$$

For compact manifolds a partition of unity consists of finitely many domains $U_\alpha$ and corresponding functions $\varphi_\alpha$.

**Theorem 5.5.** *Let $M^n$ be a compact smooth manifold. Then on $M^n$ there is a partition of unity.*

**Proof.** Take a smooth function $\psi(r)$ defined on the half-line $r \ge 0$ such that:

1) $\psi(r) \equiv 1$ for $r \le 1/2$;

2) $\psi(r) \equiv 0$ for $r \ge 1$;

3) $0 < \psi < 1$ for $1/2 < r < 1$.

Cover $M^n$ by an atlas of charts "with thick boundary". For each chart $U_\alpha$, $\alpha = 1, \dots, k$, define the function

$$\tilde{\varphi}_\alpha(x^1_\alpha, \dots, x^n_\alpha) = \psi(r) \qquad \text{for} \quad r = \sqrt{(x^1_\alpha)^2 + \dots + (x^n_\alpha)^2}.$$

The functions

$$\varphi_\alpha(x) = \frac{\tilde{\varphi}_\alpha(x)}{\sum_{\alpha=1}^k \tilde{\varphi}_\alpha(x)}$$

and the coordinate domains $U_\alpha$ form a partition of unity. ☐

By means of an atlas of charts "with thick boundary" one can easily construct an imbedding of any compact manifold into a Euclidean space of sufficiently large dimension.

**Theorem 5.6.** *Let $M^n$ be a compact smooth manifold without boundary. Then there exists its imbedding*

$$f\colon M^n \to \mathbb{R}^N$$

*into a Euclidean space of sufficiently large dimension.*

**Proof.** As in the proof of the previous theorem, take an atlas of charts "with thick boundary" $U_\alpha$, $\alpha = 1, \dots, k$, and construct the corresponding functions $\varphi_\alpha$ such that $\varphi_\alpha(x) \equiv 0$ outside $U_\alpha$, $\varphi_\alpha(x) \equiv 1$ for $r \le 1/2$, and $0 < \varphi_\alpha(x) < 1$ for $1/2 < r < 1$, where $r = |x_\alpha|$.

Define the smooth mappings

$$f_\alpha\colon M^n \to \mathbb{R}^{n+1}, \quad \alpha = 1, \dots, k,$$

by the rule

$$f_\alpha(x) = \big(\varphi_\alpha(x)x_\alpha^1, \dots, \varphi_\alpha(x)x_\alpha^n, \varphi_\alpha(x)\big) \quad \text{for} \quad x \in U_\alpha$$

and $f_\alpha(x) = 0$ for $x \in M^n \setminus U_\alpha$.

Each of these mappings specifies an imbedding of the ball $W_\alpha$ into $\mathbb{R}^{n+1}$ (see the definition of an atlas "with thick boundary"); therefore, at the points of this domain it has maximal rank equal to $n+1$. Hence the direct product of these mappings

$$f(x) = (f_1(x), \dots, f_k(x))$$

is an immersion of $M^n$ into $\mathbb{R}^{k(n+1)}$. In fact, it is an imbedding. Indeed,

a) if two points $x_1$ and $x_2$ lie in the same domain $W_\alpha$, then their images under the mapping $f_\alpha$ are different because they have different coordinates in this domain;

b) if $x_1$ lies in a domain $W_\alpha$ and $x_2$ is outside this domain, then the $(n+1)$th coordinate of $f_\alpha$ is equal to one for $x_1$ and is less than one for $x_2$.

This proves the theorem. □

The ideology of transversality developed by Whitney makes it possible to obtain imbeddings into spaces of smaller dimension by means of projections onto linear subspaces in $\mathbb{R}^N$. He has shown that:

*An imbedding of an $n$-dimensional manifold into $\mathbb{R}^N$ remains an imbedding under almost all projections $\mathbb{R}^N \to \mathbb{R}^{N-1}$ for $N > 2n+1$, and becomes an immersion for $N > 2n$.*

This is clear for curves in $\mathbb{R}^3$. The following result is a corollary.

**Theorem 5.7.** *Any compact smooth n-dimensional manifold $M^n$ can be immersed into $\mathbb{R}^{2n}$ and imbedded (realized as an n-dimensional surface) in $\mathbb{R}^{2n+1}$.*

Actually the theorem is valid also for noncompact manifolds; it is essential for the proof that the topology on $M^n$ is specified by a smooth atlas containing at most countably many charts, and $M^n$ is a Hausdorff space. Under an imbedding, the space $M^n$ inherits the metric of $\mathbb{R}^N$ and becomes a metric space. This implies that the two definitions of a smooth manifold — by means of an atlas (Section 5.1.2) and as a multidimensional surface in $\mathbb{R}^N$ (Section 5.1.2) — are equivalent.

**5.1.6. Discrete actions and quotient manifolds.** Here we point out one more method of constructing manifolds.

Let $X$ be a set and $G$ a group. The group $G$ *acts* (*from the left*) on the set $X$ if to each element $g \in G$ corresponds a one-to-one invertible transformation $\alpha(g)\colon X \to X$ satisfying, for all elements $g, h \in G$, the following conditions:

1) $\alpha(gh) = \alpha(g)\alpha(h)$;

2) $\alpha(g^{-1}) = (\alpha(g))^{-1}$;

3) $\alpha(1)$ is the identity transformation (where 1 is the unity element of $G$).

The action is *right* (the group acts from the right) if, instead of 1),

1′) $\alpha(gh) = \alpha(h)\alpha(g)$.

The action is said to be *free* if no transformation $g\colon X \to X$, except for the identity $\alpha(1)$, has fixed elements.

To simplify the notation, we will write $g$ for $\alpha(g)$ and $gx$ for $\alpha(g)(x)$.

Let $x \in X$. Then the set of all elements of the form $gx$, where $g \in G$, is called the *orbit* of the element $x$.

EXAMPLE. Let $G$ be a group and $H$ its subgroup. Then $H$ acts on $G$ by left and right translations:

$$g \to hg \text{ (left action)}, \quad g \to gh \text{ (right action)}, \quad g \in G,\ h \in H.$$

These actions are free. Orbits are the cosets (left and right, respectively) of $G$ by the subgroup $H$.

Let $X$ be a topological space, and let a group $G$ act on it by homomorphisms $g\colon X \to X$. The action is said to be *discrete* if it satisfies the following two conditions:

1) For any pair of points $x$ and $y$ with different orbits there are neighborhoods $U$ and $V$ of these points such that the orbits of these neighborhoods are disjoint.

2) For every point $x$ there is a neighborhood $U$ such that the domains $U$ and $gU$ have a nonempty intersection only for $gx = x$.

For Hausdorff spaces discreteness is provided by the following requirement: any two different points $x$ and $y$ have neighborhoods $U$ and $V$ such that the domains $gU$ and $V$ have a nonempty intersection only for finitely many elements $g \in G$. This property is often taken for definition.

**Theorem 5.8.** *Let a group $G$ act on an $n$-dimensional manifold $M^n$ by diffeomorphisms $g\colon M^n \to M^n$, which are free and discrete. Then one can introduce a smooth atlas on the set of orbits $M^n/G$ such that:*

1) *$M^n/G$ is a smooth $n$-dimensional manifold;*

2) *the projection*

$$\pi\colon M^n \to M^n/G$$

*that associates with each point of $M^n$ its orbit is a smooth mapping.*

**Proof.** Let $x \in M^n$ and let $U$ be a neighborhood of $x$ such that the domains $gU$ are pairwise disjoint for $g \in G$. We may assume that in the domain $U$ there are coordinates $x^1, \dots, x^n$. To each point of this domain corresponds exactly one orbit of action of the group $G$. In this case we say that $U$ is a chart on $M^n/G$ with coordinates $x^1, \dots, x^n$. It is specified as a projection of a domain in $M^n$. Such charts cover the entire set $M^n/G$.

Let an orbit $Gx$ lie in the intersection of two such charts, $Gx \in U_\alpha \cap U_\beta$. If these charts are obtained from neighborhoods of the point $gx$, then the change of coordinates on their intersection is given by the same functions as on the manifold $M^n$. Let $U_\alpha$ be the projection of a neighborhood of the point $g_\alpha x$, and $U_\beta$ the projection of a neighborhood of $g_\beta x$. Then the formulas of the coordinate change are given by the diffeomorphism $g_\beta g_\alpha^{-1}\colon U_\alpha \to U_\beta$ written in local coordinates as $x^i_\beta = g_\beta g_\alpha^{-1}(x^1_\alpha, \dots, x^n_\alpha)$, $i = 1, \dots, n$. Thus we obtain a smooth atlas on the space $M^n/G$. Condition 1) in the definition of discrete action implies that $M^n/G$ is a Hausdorff space. The proof is completed. □

EXAMPLE 1. On the spheres $S^n$ the group $\mathbb{Z}_2$ acts by interchanging opposite points, $x \to -x$. The quotient spaces are real projective spaces

$$\mathbb{R}P^n = S^n/\mathbb{Z}_2.$$

**Theorem 5.9.** *If the quotient manifold $M^n/G$ is orientable, then the manifold $M^n$ is also orientable and the action of the group $G$ preserves the orientation on $M^n$.*

**Proof.** Let an orientation of the manifold $M^n/G$ be chosen. We will assign positive orientation to a basis $e_1, \dots, e_n$ in the tangent space $T_xM^n$ at a point

$x \in M^n$ if the basis $\pi_*(e_1), \dots, \pi_*(e_n)$ is positively oriented. Obviously, this way we specify an orientation on $M^n$, which is preserved under the action of the group $G$. □

**Corollary 5.3.** *A projective space $\mathbb{R}P^n$ is orientable if its dimension is odd, and it is not orientable if its dimension is even.*

**Proof.** Let $S^n$ be the unit sphere in $\mathbb{R}^{n+1}$. It can be easily verified that the reflection $x \to -x$ preserves orientation if and only if the dimension of the sphere is odd. This reflection induces the action of the group $\mathbb{Z}_2$, and $\mathbb{R}P^n = S^n/\mathbb{Z}_2$. Therefore, for even $n$ the manifold $\mathbb{R}P^n$ is not orientable. For odd $n$ we will regard a basis in the tangent space at a point of $\mathbb{R}P^n$ as positively oriented if it is the image of a positively oriented basis under the mapping $\pi_*$. This orientation is well defined because in this case orientation is preserved under the reflection $x \to -x$. □

EXAMPLE 2. For any Euclidean space $\mathbb{R}^n$, the group of translations $\mathbb{Z}^n$ acts on this space:

$$(x^1, \dots, x^n) \to (x^1 + k^1, \dots, x^n + k^n), \quad (k^1, \dots, k^n) \in \mathbb{Z}^n.$$

The quotient space $T^n = \mathbb{R}^n/\mathbb{Z}^n$ is the $n$-dimensional torus.

**5.1.7. Complex manifolds.** Suppose that a smooth manifold $M^{2n}$ of even dimension is covered by an atlas of charts $\{U_\alpha\}$ with coordinates $(x^1_\alpha, \dots, x^n_\alpha, y^1_\alpha, \dots, y^n_\alpha)$ and transition functions

$$\begin{aligned} x^k_\alpha &= x^k_\alpha(x^1_\beta, \dots, x^n_\beta, y^1_\beta, \dots, y^n_\beta), \\ y^k_\alpha &= y^k_\alpha(x^1_\beta, \dots, x^n_\beta, y^1_\beta, \dots, y^n_\beta), \quad k = 1, \dots, n, \end{aligned}$$

on all intersections $U_\alpha \cap U_\beta$. Let these transition functions be written as

$$x^k_\alpha + iy^k_\alpha = z^k_\alpha = z^k_\alpha(z^1_\beta, \dots, z^n_\beta).$$

We assume that for all intersections $U_\alpha \cap U_\beta$, the transition functions are complex-analytic (holomorphic) functions of complex coordinates $z^1_\beta, \dots, z^n_\beta$,

$$\frac{\partial z^j_\alpha}{\partial \bar{z}^k_\beta} = 0, \quad j, k = 1, \dots, n,$$

with nonzero Jacobians

$$J_{\mathbb{C}} = \det\left(\frac{\partial z^j_\alpha}{\partial z^k_\beta}\right) \neq 0.$$

In this case we say that this atlas specifies a *complex* (or, in other words, *complex-analytic*) *structure* on the manifold $M^{2n}$, and the manifold is then

said to be *complex*. Locally it has the structure of the space $\mathbb{C}^n$; hence we say that its *complex dimension* $\dim_{\mathbb{C}}$ is $n$,

$$\dim_{\mathbb{C}} M^{2n} = n.$$

Introduction of a complex-analytic structure enables us to define the notion of a complex-analytic (holomorphic) function on the manifold and, in general, the notion of a *holomorphic mapping*. Namely, a smooth mapping $f\colon M^{2n} \to N^{2k}$ of smooth manifolds is said to be holomorphic if in local complex coordinates it is specified by holomorphic functions

$$w_\alpha^j = w_\alpha^j(z_\beta^1, \dots, z_\beta^n), \quad j = 1, \dots, k.$$

In particular, a mapping $f\colon M^{2n} \to \mathbb{C}^1$ is a holomorphic function on the complex manifold $M^{2n}$.

A holomorphic mapping $f\colon M \to N$ of complex manifolds is said to be *biholomorphic* if it is a diffeomorphism and the inverse mapping $f^{-1}\colon N \to M$ is also a holomorphic mapping. If there exists such a mapping between manifolds $M$ and $N$, then these manifolds are said to be *biholomorphically equivalent* or *complex-diffeomorphic*.

It is seen from the definition of a complex manifold that it must have an even dimension. This condition is not sufficient, however, for the existence of a complex structure on a manifold, as the following theorem shows.

**Theorem 5.10.** *Complex manifolds are orientable.*

**Proof.** Let $z_\alpha^j = x_\alpha^j + iy_\alpha^j$, $j = 1, \dots, n$, be complex coordinates on a manifold $M^{2n}$. In each domain $U_\alpha$ these local coordinates $(x_\alpha^1, \dots, x_\alpha^n, y_\alpha^1, \dots, y_\alpha^n)$ specify an orientation. We will show that on the intersections of charts $U_\alpha \cap U_\beta$ these orientations coincide. Indeed, by Lemma 4.2 the Jacobians of real and complex changes of coordinates are related by the formula

$$\det \begin{pmatrix} \dfrac{\partial x_\alpha^j}{\partial x_\beta^k} & \dfrac{\partial x_\alpha^j}{\partial y_\beta^k} \\[2ex] \dfrac{\partial y_\alpha^j}{\partial x_\beta^k} & \dfrac{\partial y_\alpha^j}{\partial y_\beta^k} \end{pmatrix} = J_{\mathbb{R}} = |J_{\mathbb{C}}|^2 > 0,$$

where $J_{\mathbb{C}} = \det\left(\frac{\partial z_\alpha^j}{\partial z_\beta^k}\right)$. Hence the theorem. □

This theorem and Corollary 5.3 imply the following result.

**Corollary 5.4.** *For any $n$, it is impossible to introduce a complex structure on a real projective space $\mathbb{R}P^{2n}$ of even dimension.*

Note that for a manifold of an even dimension, its orientability is still insufficient for the presence of a complex structure. At the same time, there are many examples of complex manifolds.

EXAMPLE 1. TWO-DIMENSIONAL SPHERE $S^2$. By Theorem 4.3 the two-dimensional unit sphere $(x^1)^2 + (x^2)^2 + (x^3)^2 = 1$ in $\mathbb{R}^3$ is covered by the domains $U_+$ and $U_-$, which admit diffeomorphic mappings onto the plane $x^3 = 0$ by projections from the "North" and "South Poles", $(0,0,1)$ and $(0,0,-1)$, respectively. These mappings associate complex coordinates $z_\pm$ in $U_\pm$ with complex parameters $z = x^1 \pm ix^2$ on the plane $x^3 = 0$. On the intersection of the domains $U_+ \cap U_-$ these coordinates are related by the complex-analytic mapping

$$z_+ = \frac{1}{z_-}.$$

Thus the atlas $\{(U_\pm, z_\pm)\}$ specifies a complex structure on the two-dimensional sphere $S^2$. In this way we obtain the *Riemann sphere*, which is the complex plane $\mathbb{C}$ completed with the point at infinity $z = \infty$ with local parameter $w = \frac{1}{z}$ in its neighborhood. Note that there are no complex structures on the even-dimensional spheres $S^4$ and $S^{2n}$ for $2n \geq 8$, and the question whether there exits a complex structure on the 6-dimensional sphere $S^6$ remains open.

EXAMPLE 2. COMPLEX PROJECTIVE SPACES $\mathbb{C}P^n$. A one-dimensional complex linear subspace $L \subset \mathbb{C}^{n+1}$ is specified by the formulas

$$w^1 = z_1 t, \ \ldots, \ w^{n+1} = z_{n+1} t,$$

where $(w^1, \ldots, w^{n+1})$ are complex linear coordinates in $\mathbb{C}^{n+1}$, $t$ is a parameter running over $\mathbb{C}$, and $z = (z_1, \ldots, z_{n+1})$ is the "directing vector" defined up to a constant factor. One-dimensional subspaces $L \subset \mathbb{C}^{n+1}$ form a manifold, which is the $n$-dimensional *complex projective space* $\mathbb{C}P^n$. This manifold can be defined by covering it by an atlas in the following way. Let $L_z$ be the subspace specified by a vector $z = (z_1, \ldots, z_{n+1})$ and let $z_k \neq 0$. Since this vector is determined by the subspace up to a nonzero factor, we normalize it by the condition $z_k = 1$. In the set $U_k$ of such subspaces we can introduce local complex coordinates

$$u_k^1 = z_1, \ldots, u_k^{k-1} = z_{k-1}, u_k^k = z_{k+1}, \ldots, u_k^n = z_{n+1}.$$

We take these sets $U_k$, $k = 1, \ldots, n+1$, to be the charts with local coordinates $u_k^j$. It is easily seen that the changes of coordinates on the intersections of charts $U_k \cap U_l$ are given by complex-analytic functions. For example, for $n = 1$ we obtain the *complex projective straight line* $\mathbb{C}P^1$, and $u_1 = \frac{1}{u_2}$. Thus we have proved the following fact.

**Lemma 5.4.** *The complex projective line $\mathbb{C}P^1$ and the Riemann sphere $S^2$ are biholomorphically equivalent.*

By analogy with the real case, the space $\mathbb{C}P^2$ is called the *complex projective plane*. The coordinates $(z_1, \ldots, z_{n+1})$ are called the *homogeneous*

*coordinates* on the projective space, and they are written as

$$z = (z_1 : \cdots : z_{n+1}) \in \mathbb{C}P^n$$

to express the fact that proportional vectors specify the same point of the projective space.

In a similar way the homogeneous coordinates are introduced on real projective spaces.

EXAMPLE 3. MULTIDIMENSIONAL COMPLEX SURFACES (SUBMANIFOLDS) IN $\mathbb{C}^n$. Let $f_1, \dots, f_{n-k}$ be holomorphic (complex-analytic) functions of $n$ complex variables $z^1, \dots, z^n$. Suppose that everywhere on the set of points $M$ specified by the equations

$$f_1(z^1, \dots, z^n) = \cdots = f_{n-k}(z^1, \dots, z^n) = 0$$

the matrix $\left(\frac{\partial f_l}{\partial z^m}\right)$ has the maximal rank equal to $n - k$. Then the complex version of the implicit function theorem (see Lemma 4.3) implies that $M$ is a $k$-dimensional complex manifold which, in a neighborhood of each of its points, is the graph of a holomorphic mapping. For example, if we assume without loss of generality that

$$\det\left(\frac{\partial f_l}{\partial z^m}\right) \neq 0 \quad \text{for} \quad 1 \leq l \leq n-k, \; k+1 \leq m \leq n,$$

at some point, then in a neighborhood of this point the surface $M$ is specified by

$$z^{k+1} = z^{k+1}(z^1, \dots, z^k), \quad \dots, \quad z^n = z^n(z^1, \dots, z^k).$$

In the complex case there is no analog of Whitney's theorem: multidimensional complex surfaces in complex linear spaces do not exhaust all possible complex manifolds. This is a corollary of the following theorem.

**Theorem 5.11.** *Let $f\colon M \to \mathbb{C}$ be a holomorphic function on a connected compact complex manifold $M$. Then $f = \text{const}$.*

The *proof* of this theorem will consist of several steps. First, we prove the following *maximum principle*.

**Lemma 5.5.** *Let $f\colon U \to \mathbb{C}$ be a holomorphic function defined on a connected domain in $\mathbb{C}^n$. Then if its modulus $|f|$ attains its maximal value at an interior point of $U$, then $f$ is a constant. In particular, if the domain $U$ is bounded, then $|f|$ attains its maximum on the boundary of $U$.*

**Proof.** Let $z_0$ be a local maximum point of $|f|$ interior to $U$. If we restrict the function $|f|$ to any complex straight line passing through $z_0$, then the point $z_0$ will also be a local maximum point for the restriction. Hence we may prove the lemma for the case $n = 1$. Without loss of generality let

$z_0 = 0$. We normalize the function $f$ so that its value at zero be real and nonnegative,

$$f(0) \geq 0, \qquad f(0) \in \mathbb{R}.$$

Cauchy's integral formula, which is one of the corollaries of the Stokes theorem (see Chapter 9), says that

$$f(0) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)\,dz}{z},$$

where $\gamma$ is a contour enclosing the origin. Take for $\gamma$ the circle $z = re^{i\varphi}$ of a small radius $r$. Then

$$\begin{aligned} f(0) &= \frac{1}{2\pi i} \int_\gamma \frac{f(re^{i\varphi})\,d(re^{i\varphi})}{re^{i\varphi}} \\ &= \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(re^{i\varphi}) ire^{i\varphi}\,d\varphi}{re^{i\varphi}} = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\varphi})\,d\varphi. \end{aligned}$$

This formula implies that

$$\int_0^{2\pi} \operatorname{Re}\left(f(0) - f(re^{i\varphi})\right) d\varphi = 0,$$

and, since $f(0) = |f(0)| \geq |f(re^{i\varphi})| \geq \operatorname{Re} f(re^{i\varphi})$ by the local maximum condition, this equality is possible only for $f(0) \equiv f(re^{i\varphi}) = \operatorname{Re} f(re^{i\varphi})$. This holds for all sufficiently small $r$. Therefore, the holomorphic function $f(z)$ is constant in a neighborhood of the local maximum point $z = 0$, and hence it is constant in any connected domain containing this point. □

**Proof of Theorem 5.11.** Let $f\colon M \to \mathbb{C}$ be a holomorphic function on a connected compact manifold. Due to compactness, $|f|$ attains its maximum at some point $z_0 \in M$. The manifold $M$ is locally like $\mathbb{C}^n$; hence in some coordinate neighborhood of $z_0$ the function $f$ equals a constant $C$. Obviously, the set $N$ specified by the equation $f = C$ is closed. Any point of it is a maximum point of $|f|$; hence it is contained in $N$ together with some neighborhood. Therefore, the set $N$ is open and nonempty. It remains to note that a nonempty open and closed subset $N$ of a connected topological space $M$ coincides with the entire $M$. □

**Corollary 5.5.** *Multidimensional complex surfaces in $\mathbb{C}^n$ are not compact.*

The proof is simple: each coordinate $z^j$, $j = 1, \ldots, n$, on the space $\mathbb{C}^n$ is a holomorphic function on a connected surface, and if the surface were compact, then by the preceding theorem the coordinate $z^j$ would have to be constant.

EXAMPLE 4. RIEMANN SURFACES. Consider a one-dimensional complex submanifold in $\mathbb{C}^2$ specified by an equation

$$f(z,w)=0.$$

We assume that the complex gradient does not vanish at any point of the surface,

$$\operatorname{grad}_{\mathbb{C}} f=\left(\frac{\partial f}{\partial z},\frac{\partial f}{\partial w}\right)\neq 0.$$

Therefore, in a neighborhood of each point the equation $f(z,w)=0$ can be solved for one of the variables. For example, if $\frac{\partial f}{\partial w}\neq 0$ at a given point, we obtain that the surface is the graph of a function $w=w(z)$. By analytic continuation of this relation to the entire complex plane $\mathbb{C}$ we obtain a representation of the submanifold as the graph of a function which may be multivalued. Let us give some examples.

a) If $f(z,w)=z-e^w$, we obtain the graph of the multivalued logarithmic function,

$$w=\log z=\log|z|+i\arg z+2\pi ik,\quad k=1,2,\dots,$$

defined for $z\neq 0$. In a sufficiently small neighborhood of each point of the surface, the projection $(z,w)\to z$ onto the $z$-plane is a diffeomorphism of this neighborhood onto its image.

b) If $f(z,w)=w^2-F(z)$, where the function $F$ and its derivative $F_z$ with respect to $z$ do not simultaneously vanish at the same point, we have a *hyperelliptic surface* specified by the formula

$$w=\sqrt{F(z)}.$$

The projection $(z,w)\to z$ is a diffeomorphism in a neighborhood of any point such that $w\neq 0$.

A surface in $\mathbb{C}^2$ specified by an equation $f(z,w)=0$, where $f$ is a complex-analytic function, is called a *Riemann surface.* It is represented as the graph of a multivalued complex-analytic function

$$w=w(z).$$

The points of a Riemann surface at which the Jacobian of the projection $(z,w)\to z$ vanishes, are called the *branch points* of the surface. This condition is equivalent to

$$\frac{\partial f}{\partial w}=0.$$

In some cases a Riemann surface may be made a compact complex manifold by adding finitely many points. Consider a hyperelliptic Riemann surface

$$w=\sqrt{(z-z_1)\cdots(z-z_n)}$$

branching at finitely many points $z_1,\dots,z_n$.

Assume first that the number of these points is even, $n = 2g+2$, $g \geq 0$. In this case we delete from the $z$-plane the pairwise disjoint segments $\gamma_j$, $j = 1, \dots, g+1$, which are curves joining the consecutive pairs of branch points, $z_1$ and $z_2$, $z_3$ and $z_4$, $\dots$, $z_{2g+1}$ and $z_{2g+2}$. Let $\Gamma$ be the $z$-plane with deleted segments $\gamma_j$, $j = 1, \dots, g+1$. On the plane $\Gamma$ the function

$$w = \sqrt{(z - z_1) \cdots (z - z_{2g+2})}$$

is two-valued, and its graph falls into two copies $\Gamma_+$ and $\Gamma_-$ of a surface that is complex-diffeomorphic to $\Gamma$. These surfaces are called the *branches* of the multivalued function $w = w(z)$.

Note that $\Gamma$ may be supplemented by the point at infinity to become the Riemann sphere $S^2 = \mathbb{C}P^1$ with deleted segments $\gamma_j$, $j = 1, \dots, g+1$. Denote this complex manifold also by $\Gamma$. The graph of the function $w$ is also completed to two copies, $\Gamma_\pm$, of $\Gamma$. In a neighborhood of the point at infinity $z = \infty$, we will have then two different branches of $w$ with asymptotics

$$w = \pm z^{g+1}\left(1 + O\left(\frac{1}{z}\right)\right)$$

on $\Gamma_\pm$. The projection $\Gamma_+ \cup \Gamma_- \to \Gamma$ of the graph of $w$ onto the $z$-plane given by $(z, w) \to z$ is a holomorphic mapping. In this way we have completed a $\mathbb{C}$-valued holomorphic function to a holomorphic mapping into the Riemann sphere $\mathbb{C}P^1$.

If a curve $\gamma$ intersects the segment $\gamma_j$, then its image $w(\gamma)$ will be cut by the intersection point into two pieces lying near this point on different branches of the function $w(z)$. A segment deleted from $\Gamma$ may be approached from opposite directions, let us say, from the left and from the right. Riemann constructed a new surface by identifying for each $j = 1, \dots, g+1$ the left-hand side of the curve $\gamma_j$ on $\Gamma_+$ with its right-hand side on $\Gamma_-$ and the right-hand side of $\gamma_j$ on $\Gamma_+$ with its left-hand side on $\Gamma_-$. The result, the Riemann surface of the function $w(z)$, is a complex surface on which the holomorphic mapping $w$ into $\mathbb{C}P^1$ is single-valued.

This surface has the following topological structure. Recall that $\Gamma_\pm$ are two copies of the sphere $S^2$ with $g+1$ deleted segments. Such a sphere is homeomorphic to the sphere $S^2$ with $g+1$ closed disks deleted. Identifying different sides of the segments $\gamma_j$ on the two spheres is equivalent, from the point of view of topology, to the following operation: on each of the two spheres we take the boundary of the deleted disk and glue the cylinder $S^1 \times [0,1]$ to these two circles. We leave it as an exercise for the reader to show that *the Riemann surface of the function* $w = \sqrt{(z - z_1) \cdots (z - z_{2g+2})}$ *is homeomorphic to the connected sum of* $g$ *tori*, which is a *sphere with* $g$ *handles.*

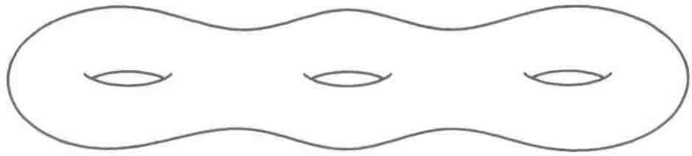

**Figure 5.8.** A sphere with three handles.

The function

$$w = \sqrt{(z - z_1) \cdots (z - z_{2g+1})}$$

with an odd number of finite branch points does not allow for exactly the same construction. In this case one of the segments $\gamma_{g+1}$ must join $z_{2g+1}$ with the point at infinity. On the Riemann sphere, there is no difference between finite points and the point $\infty$, hence such an analog is well defined. Then, identifying the left- and right-hand sides of boundary segments of different branches, we construct the Riemann surface on which $w(z)$ is single-valued. It may be proved similarly to the previous case that

*The Riemann surface of the function* $w = \sqrt{(z - z_1) \cdots (z - z_{2g+1})}$ *is homeomorphic to the connected sum of $g$ tori, which is a sphere with $g$ handles.*

A compact Riemann surface may be constructed for any surface of the form $f(z, w) = 0$, where $f$ is a polynomial. The added points correspond to different asymptotics of the multivalued function $w = w(z)$ at infinity. Any compact Riemann surface $\Sigma$ is diffeomorphic to the connected sum of finitely many tori. The number of these tori is called the *genus of the Riemann surface* $\Sigma$.

The Riemann surfaces corresponding to polynomials $f(z, w)$ admit the following immersion into $\mathbb{C}P^2$. Making the substitution

$$z = \frac{z_1}{z_3}, \qquad w = \frac{z_2}{z_3}$$

in the polynomial $f(z, w)$ and multiplying it by $z_3^D$, where $D$ is its degree, we obtain an equation of the form

$$P(z_1, z_2, z_3) = 0$$

written in homogeneous coordinates $(z_1 : z_2 : z_3)$ in $\mathbb{C}P^2$. The points at infinity correspond to the solutions of this equation with $z_3 = 0$. They may glue to each other, and the "finite" part of the Riemann surface $f(z, w) = 0$ defined by the condition $z_3 \neq 0$ will be imbedded into $\mathbb{C}P^2$.

For example, the equation

$$w^2 = z^3 - 1,$$

with the corresponding equation

$$z_2^2 z_3 = z_1^3 - z_3^3$$

in homogeneous coordinates, specifies a surface of genus one. The projection $\mathbb{C}P^1 \to \mathbb{C}P^1$, as shown above, has three (finite) branch points on the $z$-plane, which are the roots of the equation $z^3 - 1 = 0$. The point $z_3 = 0$ is a double point, at which two compactified branches of the function $\sqrt{z^3 - 1}$ intersect under the immersion into $\mathbb{C}P^2$.

EXAMPLE 5. ALGEBRAIC MANIFOLDS. We say that $M$ is a *complex submanifold* of a complex manifold $N$ if it can be imbedded into $N$ as a smooth submanifold,

$$f: M \to N,$$

and the imbedding is specified by holomorphic (complex-analytic) functions. Like all smooth manifolds, $M$ is locally determined by equations

$$f_1(z^1, \dots, z^n) = \cdots = f_k(z^1, \dots, z^n) = 0,$$

but in this case $f_1, \dots, f_k$ are holomorphic functions of complex coordinates in $N$.

The complex submanifolds of complex projective spaces are called *algebraic manifolds.*

By the Chow theorem, if $M$ is a complex submanifold of the space $\mathbb{C}P^n$ and $\dim_{\mathbb{C}} M = k$, then $M$ is the set of zeros of a family of homogeneous polynomials of homogeneous coordinates $(z_1 : \cdots : z_{n+1})$ in $\mathbb{C}P^n$,

$$P_1(z_1, \dots, z_{n+1}) = \cdots = P_{n-k}(z_1, \dots, z_{n+1}) = 0.$$

This is the reason that such manifolds are called algebraic. Recall that a function $f(z_1, \dots, z_{n+1})$ is said to be *homogeneous* if for any constant $\lambda$,

$$f(\lambda z_1, \dots, \lambda z_{n+1}) = \lambda^D f(z_1, \dots, z_{n+1}),$$

where the constant $D$ is the degree of homogeneity. The sets of zeros of homogeneous functions are invariant relative to homotheties $z \to \lambda z$.

In contrast to complex submanifolds in $\mathbb{C}^n$, we have the following

**Theorem 5.12.** *Algebraic manifolds are compact.*

**Proof.** First of all we will prove that complex projective spaces $\mathbb{C}P^n$ are compact. The equation

$$|z_1|^2 + \cdots + |z_{n+1}|^2 = 1 \tag{5.7}$$

specifies the $(2n+1)$-dimensional sphere in $\mathbb{C}^{n+1} = \mathbb{R}^{2n+2}$. The continuous mapping

$$f: S^{2n+1} \to \mathbb{C}P^n, \quad f(z_1, \dots, z_{n+1}) = (z_1 : \cdots : z_{n+1}) \in \mathbb{C}P^n,$$

assigns to each point of the sphere a point in $\mathbb{C}P^n$. The homogeneous coordinates of any point in $\mathbb{C}P^n$ may be multiplied by a constant to make them satisfy (5.7). Thus $f(S^{2n+1}) = \mathbb{C}P^n$, and Lemma 5.1 implies that the space $\mathbb{C}P^n$ is also compact being the image of a compact set under a continuous mapping.

Algebraic manifolds are closed subspaces in $\mathbb{C}P^n$ and by Lemma 5.1 are also compact. □

EXAMPLE. QUADRICS IN $\mathbb{C}P^2$ AND $\mathbb{C}P^3$. The equation

$$z_1^2 + z_2^2 + z_3^2 = 0$$

specifies the Riemann sphere imbedded into $\mathbb{C}P^2$. Indeed, let $z = \frac{z_1}{z_3}$, $w = i\frac{z_2}{z_3}$. Then for $z_3 \neq 0$ this equation becomes

$$z^2 = w^2 - 1.$$

The points $(1 : 1 : 0)$ and $(1 : -1 : 0)$ for which $z_3 = 0$ complete this surface in $\mathbb{C}^2$ to the Riemann sphere in $\mathbb{C}P^2$.

Consider now the equation

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 = 0 \tag{5.8}$$

specifying a complex surface in $\mathbb{C}P^3$.

Let us introduce new homogeneous coordinates

$$z_1' = z_1 + iz_2, \quad z_2' = z_1 - iz_2, \quad z_3' = i(z_3 + iz_4), \quad z_4' = i(z_3 - iz_4).$$

In terms of these coordinates the equation (5.8) becomes

$$z_1'z_2' = z_3'z_4'. \tag{5.9}$$

Take two copies of the Riemann sphere $\mathbb{C}P^1$ with homogeneous coordinates $(r_1 : r_2)$ and $(s_1 : s_2)$ and define a mapping $\mathbb{C}P^1 \times \mathbb{C}P^1 \to \mathbb{C}P^3$ by the formula

$$[(r_1 : r_2), (s_1 : s_2)] \to (z_1' : z_2' : z_3' : z_4') = (r_1s_1 : r_2s_2 : r_1s_2 : r_2s_1).$$

We leave it to the reader as a simple exercise to verify that this mapping is an imbedding and its image is the surface specified by equation (5.9). Thus we have proved the following assertion.

**Theorem 5.13.** *The equation*

$$z_1^2 + \cdots + z_n^2 = 0$$

*specifies the Riemann sphere $\mathbb{C}P^1 \subset \mathbb{C}P^2$ for $n = 3$ and the direct product of two spheres, $\mathbb{C}P^1 \times \mathbb{C}P^1 \subset \mathbb{C}P^3$, for $n = 4$.*

EXAMPLE 6. COMPLEX TORI are quotient spaces of $\mathbb{C}^n$ by the action of the discrete subgroup $\Lambda \simeq \mathbb{Z}^{2n}$ of the additive group $\mathbb{C}^n$. The quotient space $\mathbb{C}^n/\mathbb{Z}^{2n}$ is diffeomorphic to the $2n$-dimensional real torus $T^{2n}$, and in a small neighborhood of each point, linear coordinates in $\mathbb{C}^n$ determine the complex coordinates on the torus. Therefore, the quotient space is a complex manifold. It is called the $n$-dimensional *complex torus.*

**Theorem 5.14.** *Each one-dimensional complex torus is complex-diffeomorphic to a torus of the form* $E_\tau = \mathbb{C}/\{m+n\tau : m, n \in \mathbb{Z}\}$, *where* $\operatorname{Im}\tau > 0$; *two tori* $E_\tau$ *and* $E_{\tau'}$ *are complex-diffeomorphic if and only if the parameters* $\tau$ *and* $\tau'$ *are related by a linear-fractional transformation,*

$$\tau' = \frac{a\tau + b}{c\tau + d},$$

*where* $a, b, c, d \in \mathbb{Z}$ *and* $ad - bc = \det\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = 1$.

**Proof.** Let $T = \mathbb{C}/\Lambda$ be a one-dimensional complex torus, and let $e_1, e_2$ be a basis of the lattice $\Lambda$. Replacing possibly $e_2$ by $-e_2$, we reduce the basis to the form where $\operatorname{Im}\frac{e_2}{e_1} > 0$ and let $\tau = e_2/e_1$. Take the vector $e_1$ as the basis of the complex linear space $\mathbb{C}$ with coordinate $z$. We see that in these coordinates the torus has the form $T = \mathbb{C}/\{m + n\tau : m, n \in \mathbb{Z}\}$.

The linear transformation of the real space $\mathbb{R}^2 = \mathbb{C}$ specified by a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

maps the subgroup $\Lambda$ into itself. This implies that the subgroups generated by the vectors $(1, \tau)$ and $(c\tau + d, a\tau + b)$ coincide. Therefore, the tori $E_\tau$ and $E_{\tau'}$ coincide if $\tau$ and $\tau'$ are related by the corresponding linear-fractional transformation. We omit the proof of the converse statement. □

**Corollary 5.6.** *There exist compact complex manifolds that are diffeomorphic, but not complex-diffeomorphic.*

If a complex torus is an algebraic manifold, then it is said to be *Abelian.* The following is the well-known Riemann theorem.

*A complex torus* $\mathbb{C}^n/\Lambda$ *is Abelian if and only if, in appropriate linear coordinates in* $\mathbb{C}^n$, *it reduces to the form*

$$\mathbb{C}^n/\{\Delta M + BN : M, N \in \mathbb{C}^n\},$$

*where* $\Delta$ *is a diagonal matrix with positive integer diagonal elements, and* $B$ *is a symmetric matrix such that its imaginary part is positive definite,*

$$B_{jk} = B_{kj}, \quad \operatorname{Im} B_{jk}\eta^j\eta^k > 0 \quad \text{for} \quad \eta \in \mathbb{R}^n, \eta \neq 0.$$

**Corollary 5.7.** 1. *All tori of complex dimension* 1 *are Abelian.*

2. *For $n \geq 2$, almost all $n$-dimensional complex tori are non-Abelian.*

The meaning of this statement is as follows. For a given basis $e_1, \dots, e_{2n}$ in $\mathbb{R}^{2n} = \mathbb{C}^n$, a subgroup $\Lambda$ isomorphic to $\mathbb{Z}^{2n}$ is specified by the coordinates of the vectors $\eta_1, \dots, \eta_{2n}$ generating the subgroup $\Lambda$. These coordinates $\eta_j^k$, where $\eta_j = \eta_j^k e_k$, lie in a domain in $\mathbb{R}^{4n^2}$, and "almost all" points of this domain correspond to non-Abelian tori. "Almost all" means that the intersection of the complement of the set of these points with any ball in $\mathbb{R}^{4n^2}$ has zero measure.

If $\Delta$ is the identity matrix, then one says that the Abelian torus has *principal polarization.* In this case the famous Riemann *theta-function*

$$\theta(z_1, \dots, z_n) = \sum_{M \in \mathbb{Z}^n} \exp i\Big(\frac{1}{2} B_{jk} M^j M^k + z_j M^j\Big)$$

is defined as an entire function of $n$ complex variables $z_1, \dots, z_n$.

## 5.2. Groups of transformations as manifolds

**5.2.1. Groups of motions as multidimensional surfaces.** Denote by $\mathrm{M}(n, \mathbb{R})$ the space of all $n \times n$ matrices (or matrices of the $n$th order) with real elements. This is a linear space of dimension $n^2$, and the matrix entries $a_j^i$ are the Cartesian coordinates in $\mathrm{M}(n, \mathbb{R}) = \mathbb{R}^{n^2}$.

If $A = (a_j^i)$, $B = (b_j^i)$ are two matrices in $\mathrm{M}(n, \mathbb{R})$, then their product $C = (c_j^i)$, where

$$c_j^i = a_k^i b_j^k, \tag{5.10}$$

also belongs to $\mathrm{M}(n, \mathbb{R})$. Hence the multiplication law (5.10) specifies a smooth mapping of the direct product $\mathrm{M}(n, \mathbb{R}) \times \mathrm{M}(n, \mathbb{R})$ into $\mathrm{M}(n, \mathbb{R})$,

$$\mathrm{M}(n, \mathbb{R}) \times \mathrm{M}(n, \mathbb{R}) \to \mathrm{M}(n, \mathbb{R}),$$

with $(A, B) \to AB$.

The radius-vectors of points in $\mathrm{M}(n, \mathbb{R})$ are defined in a natural way as $n \times n$ matrices. On the corresponding vector space we define the Euclidean scalar product

$$\langle A, B \rangle = \sum_{i,j=1}^{n} a_j^i b_j^i, \quad A = (a_j^i), \quad B = (b_j^i).$$

It is clear that the matrix elements specify Euclidean coordinates in $\mathbb{R}^{n^2}$. The scalar product determines the matrix norm by the formula

$$|A|^2 = \sum_{i,j} (a_j^i)^2, \quad A = (a_j^i).$$

Obviously,

$$|A+B| \leq |A| + |B|. \tag{5.11}$$

Moreover, this norm satisfies the inequality

$$|AB| \leq |A||B|, \tag{5.12}$$

which is written in the coordinate form as

$$\sqrt{\sum_{i,j,k=1}^{n}(a_k^i)^2(b_j^k)^2} \leq \sqrt{\sum_{i,k=1}^{n}(a_k^i)^2} \cdot \sqrt{\sum_{j,k=1}^{n}(b_j^k)^2}.$$

The latter inequality is a consequence of the inequality (see 2.1.1)

$$\left(\sum x_i y_i\right)^2 \leq \left(\sum x_i^2\right)\left(\sum y_i^2\right).$$

**Lemma 5.6.** *If $|X| < 1$, then the matrix $A = 1 + X$ is invertible.*

**Proof.** Write down the series

$$B = 1 - X + X^2 - X^3 + \cdots + (-1)^n X^n + \cdots. \tag{5.13}$$

It follows from inequalities (5.11) and (5.12) that for any $k$, $l$,

$$\begin{aligned}&|X^k - X^{k+1} + \cdots + (-1)^{l-1}X^{k+l-1}|\\ &\quad \leq |X|^k(1 + |X| + \cdots + |X|^{l-1}) \leq |X|^k \frac{1-|X|^l}{1-|X|} \leq \frac{|X|^k}{1-|X|}.\end{aligned}$$

This implies that the partial sums of the series (5.13) form a Cauchy sequence for $|X| < 1$; hence this series converges. Multiplying $A$ by $B$, we obtain

$$AB = (1+X)(1 - X + X^2 - X^3 + \cdots) = 1.$$

Thus $B = A^{-1}$. □

This lemma implies that the group $\mathrm{GL}(n, \mathbb{R})$ contains a neighborhood of the identity matrix in $\mathrm{M}(n, \mathbb{R})$. In fact, this property holds for any matrix in $\mathrm{GL}(n, \mathbb{R})$. Indeed, let $A \in \mathrm{GL}(n, \mathbb{R})$. Then the matrices of the form $A(1 + X)$, where $|X| < 1$, constitute a neighborhood of $A$ in $\mathrm{M}(n, \mathbb{R})$ and all of them are invertible,

$$[A(1+X)]^{-1} = (1+X)^{-1}A^{-1}.$$

Therefore, this neighborhood lies in $\mathrm{GL}(n, \mathbb{R})$, and the elements of the matrix $X$ specify local coordinates in a neighborhood of the matrix $A \in \mathrm{GL}(n, \mathbb{R})$. Thus we have proved that the group $\mathrm{GL}(n, \mathbb{R})$ is a domain in $\mathrm{M}(n, \mathbb{R})$.

The tangent space to the group $\mathrm{GL}(n, \mathbb{R})$ at its unit element is the space of all matrices of order $n$. The dimension of this group is $n^2$.

The following lemma shows that if $G$ is a subgroup of $\mathrm{GL}(n, \mathbb{R})$ specified by a system of equations $F = 0$, then it suffices to verify its regularity as a surface in $\mathbb{R}^{n^2}$ only in a neighborhood of the identity.

**Lemma 5.7.** *Let $G \subset \mathrm{GL}(n, \mathbb{R})$ be a subgroup specified in a neighborhood $U$ of the unit element as the set of zeros of a smooth mapping*

$$F: U \to \mathbb{R}^k$$

*(i.e., by the set of common zeros for $k$ functions $F_1, \ldots, F_k = 0$), and all the points of $G$ are nonsingular. Then $G$ is an $(n^2 - k)$-dimensional regular surface in $\mathrm{M}(n, \mathbb{R})$.*

**Proof.** Let $A \in G$. Consider the neighborhood $AU$ of the matrix $A$ consisting of the matrices $AX$, where $X \in U$. Define on $AU$ a smooth mapping $F_A$:

$$F_A(AX) = F(X).$$

The equation $F_A = 0$ determines the elements of $G$ in a neighborhood of $A$, and all the points of $G$ are nonsingular. Since $A \in G$ is an arbitrary matrix, $G$ is a regular surface of dimension $(n^2 - k)$. □

The subgroup $\mathrm{SL}(n, \mathbb{R})$ of $\mathrm{GL}(n, \mathbb{R})$ is specified by the equation

$$\det A = 1. \tag{5.14}$$

**Theorem 5.15.** $\mathrm{SL}(n, \mathbb{R})$ *is an $(n^2 - 1)$-dimensional nonsingular surface in $\mathbb{R}^{n^2}$. The tangent space to $\mathrm{SL}(n, \mathbb{R})$ at the unit element is the space of all matrices with zero trace, $\mathrm{Tr}\, X = 0$.*

**Proof.** Let $A(t)$ be a smooth curve in $\mathrm{SL}(n, \mathbb{R})$ passing through the identity, $A(0) = 1$. Write down the Taylor expansion of $A(t)$,

$$A(t) = 1 + Xt + O(t^2),$$

and substitute this expansion into (5.14):

$$\det A = 1 + t\, \mathrm{Tr}\, X + O(t^2) = 1.$$

Hence the tangent vectors to $\mathrm{SL}(n, \mathbb{R})$ at the identity are the zero-trace matrices,

$$\mathrm{Tr}\, X = \sum_{i=1}^{n} x_i^i = 0, \quad X = (x_j^i).$$

Lemma 5.3 implies that the unit element of the group and the points of $\mathrm{SL}(n, \mathbb{R})$ lying close enough to it are nonsingular. By Lemma 5.7 the entire group $\mathrm{SL}(n, \mathbb{R})$ is a nonsingular surface. □

The group of orthogonal matrices $\mathrm{O}(n)$ is specified in $\mathbb{R}^{n^2}$ by the equation

$$A^\top A = 1, \tag{5.15}$$

which can be written as

$$\sum_{k=1}^{n} a_k^i a_k^j = \delta^{ij}, \quad i, j = 1, \dots, n.$$

Interchanging of $i$ and $j$ yields the same equation. Hence there are $n^2 - n(n+1)/2 = n(n-1)/2$ equations which correspond to the pairs $i = j$ and $i < j$.

**Theorem 5.16.** *The group* $\mathrm{O}(n)$ *is a nonsingular surface of dimension* $\frac{n(n-1)}{2}$. *The tangent space to* $\mathrm{O}(n)$ *at the identity consists of all skew-symmetric matrices.*

**Proof.** Let $A(t)$ be a curve in $\mathrm{O}(n)$ such that $A(0) = 1$. We have the Taylor expansion

$$A(t) = 1 + Xt + O(t^2).$$

Substituting it into the equation (5.15) yields

$$(1 + X^\top t + O(t^2))(1 + Xt + O(t^2)) = 1 + (X^\top + X)t + O(t^2) = 1.$$

Therefore, the equations (5.6) for the surface $\mathrm{O}(n)$ specify an $n(n-1)/2$-dimensional space consisting of all skew-symmetric matrices,

$$X^\top = -X, \qquad x_j^i = -x_i^j, \quad X = (x_j^i).$$

Now the theorem follows from Lemmas 5.3 and 5.7. $\square$

This argument can be extended to the group of linear transformations preserving any nondegenerate scalar products in $\mathbb{R}^n$. Namely, if $G$ is the Gram matrix of a scalar product, then it is preserved by a transformation $A \in \mathrm{GL}(n, \mathbb{R})$ if and only if

$$A^\top G A = G,$$

which implies that

$$(1 + X^\top t + O(t^2))G(1 + Xt + O(t^2)) = G + (X^\top G + GX)t + O(t^2) = G.$$

Hence we obtain the following equation for tangent vectors:

$$X^\top G + GX = 0. \tag{5.16}$$

If the scalar product is symmetric (e.g., pseudo-Euclidean), then $G^\top = G$ and relation (5.16) means that the matrix $GX$ is skew-symmetric. For a nondegenerate scalar product ($\det G \neq 0$) this condition specifies an $n(n-1)/2$-dimensional subspace. This implies the following result.

**Theorem 5.17.** *The groups are nonsingular surfaces in* $\mathbb{R}^{n^2}$ *of dimension* $\frac{n(n-1)}{2}$, *where* $n = p + q$. *Their tangent spaces at the identity consist of all matrices satisfying the equation*

$$(GX)^\top + GX = 0,$$

*where* $G$ *is a diagonal* $n \times n$ *matrix of the form*

$$G = \operatorname{diag}(\underbrace{1, \ldots, 1}_{p}, \underbrace{-1, \ldots, -1}_{q}).$$

If the scalar product is skew-symmetric (e.g., symplectic), then $G^\top = -G$ and the equality (5.16) implies that the matrix $GX$ is symmetric.

Recall that symplectic transformations of the space $\mathbb{R}^{2n}$ form the group $\mathrm{Sp}(n, \mathbb{R})$. This group consists of matrices $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, where $A$, $B$, $C$, and $D$ are $n \times n$ matrices satisfying the equation

$$\begin{pmatrix} A^\top & C^\top \\ B^\top & D^\top \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In particular, as we have shown before, $\mathrm{Sp}(1, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})$.

Similarly to the above one can prove the following assertion.

**Theorem 5.18.** *The group* $\mathrm{Sp}(n, \mathbb{R})$ *is a* $(2n^2 + n)$*-dimensional nonsingular surface in* $\mathbb{R}^{4n^2}$.

The subgroups $\mathrm{SO}(n)$ and $\mathrm{SO}(p, q)$ of the groups $\mathrm{O}(n)$ and $\mathrm{O}(p, q)$ are also nonsingular surfaces of the same dimension as the corresponding groups.

The groups $\mathrm{O}(n)$ and $\mathrm{O}(p, q)$ are not connected. This follows from the fact that the smooth function det can take the values $\pm 1$, and matrices with determinants of opposite signs cannot by joined by a curve in the group.

A submanifold $N^k \subset M^k$ is called a *connected component* of the manifold $M^k$ if it is connected and the manifold $M^k$ is a union $M^k = N^k \cup L^k$ of two disjoined submanifolds which are closed and open at the same time as subsets of $M^k$.

Before proving that $\mathrm{SO}(n)$ is a connected component of the group $\mathrm{O}(n)$ we will prove the following statement.

**Lemma 5.8.** *Let* $A \in \mathrm{SO}(n)$. *Then the* $n$*-dimensional Euclidean vector space* $\mathbb{R}^n$ *is the direct sum* $\mathbb{R}^n = V_0 \oplus V_1 \oplus \cdots \oplus V_m$ *of subspaces* $V_j$, $j \geq 0$, *such that*

1) *the* $V_j$ *are pairwise orthogonal;*

2) $A$ *is the identity transformation on* $V_0$;

3) *the* $V_j$ *are two-dimensional for* $j \geq 1$, *and* $A$ *acts on each of them as rotation through an angle* $\varphi_j$.

**Proof.** We will prove the lemma by induction on $n$. For $n = 1$ the assertion is trivial, since $A = 1$. Assume that it holds for $k < n$. Let $A \in \mathrm{SO}(n)$. Write down the characteristic equation

$$\det(A - \lambda \cdot 1) = 0.$$

If $\mu$ is a complex root of this equation, then there exists a complex vector $\eta = \xi_1 + i\xi_2$, where $\xi_1, \xi_2 \in \mathbb{R}^n$, such that $A\eta = \mu\eta$. Obviously, $A(\xi_1 - i\xi_2) = \bar{\mu}(\xi_1 - i\xi_2)$ and the vectors $\xi_1, \xi_2$ generate a two-dimensional invariant subspace $W = \mathbb{R} \cdot \xi_1 + \mathbb{R} \cdot \xi_2$. Since the operator $A$ preserves the lengths of vectors, we obtain that $|\mu| = 1$, $\mu = \cos\varphi - i\sin\varphi$, and

$$A \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}.$$

This implies that $|\xi_1| = |\xi_2|$ and $\langle \xi_1, \xi_2 \rangle = 0$. Hence we can assume that $|\xi_1| = |\xi_2| = 1$.

Now we decompose the vector space $\mathbb{R}^n$ into the direct sum of the subspace $W$ and its orthogonal complement $W^\perp$. This decomposition is invariant with respect to the action of the operator $A$: $AW = W$ and $AW^\perp = W^\perp$. It remains to apply the induction assumption to the restriction of $A$ to $W^\perp$.

If all the roots of the equation $\det(A - \lambda \cdot 1) = 0$ are real, then they are equal to $\pm 1$, and since $\det A = 1$, the number of the roots equal to $-1$ is even. If all the roots are equal to $+1$, we set $\mathbb{R}^n = V_0$. Otherwise, we take two orthogonal vectors $\xi_1$ and $\xi_2$ such that $A\xi_i = -\xi_i$, $i = 1, 2$. On the subspace $W$ spanned by them, $A$ acts as rotation through the angle $\pi$. The decomposition $\mathbb{R}^n = W + W^\perp$ is invariant, and for $W^\perp$ the lemma holds by the induction assumption. □

Now we deduce connectedness of the space $\mathrm{SO}(n)$ from this lemma.

**Theorem 5.19.** *The space* $\mathrm{SO}(n)$ *is a connected component of the group* $\mathrm{O}(n)$.

**Proof.** In order to show that $\mathrm{SO}(n)$ is connected, it suffices to show that any transformation $A \in \mathrm{SO}(n)$ can be continuously reduced inside $\mathrm{SO}(n)$ to the identity of the group. As we have shown, any element $A \in \mathrm{O}(n)$ may be represented in an appropriate basis by a block-diagonal matrix

$$\begin{pmatrix} 1 & & 0 \\ & A_1 & \\ & & \\ 0 & & A_k \end{pmatrix}, \tag{5.17}$$

where the matrices $A_j$ specify rotations through angles $\varphi_j$. By varying the $\varphi_j$ continuously, we can transform each matrix $A_j$ into the identity matrix. □

Thus we have shown that SO(3) is a connected three-dimensional surface in $\mathbb{R}^9 = \mathrm{M}(3, \mathbb{R})$. A system of local coordinates on SO(3) provided by Euler's angles is well known from analytic geometry. A rotation which transforms a coordinate system $(x, y, z)$ into $(x', y', z')$ can be represented as the composition of the following three rotations.

1. Rotation through the angle $\varphi$ about the $z$-axis. The $x$-axis turns into the line of nodes.

2. Rotation through the angle $\theta$ about the line of nodes. The $z$-axis turns into the $z'$-axis.

3. Rotation through the angle $\psi$ about the $z'$-axis. The line of nodes turns into the $x$-axis.

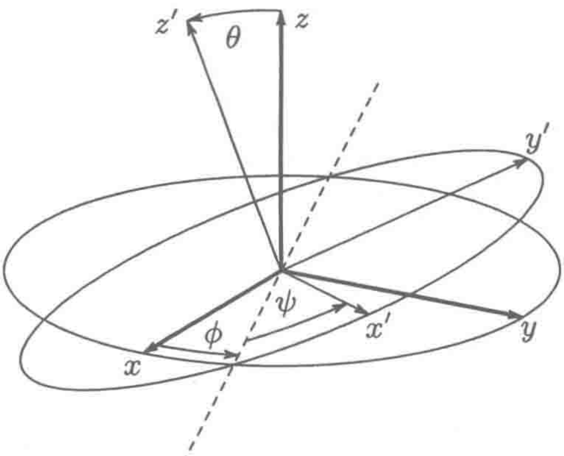


**Figure 5.9.** Euler's angles (the dotted line is the line of nodes).

From the classification of motions of $\mathbb{R}^3$ given in Section 1.3.3 we know that any transformation in SO(3) is representable as a rotation about some axis. Let us assign to such a rotation a point of the sphere $|x| \leq \pi$. The radius-vector of this point specifies the axis of rotation, and $|x|$, the rotation angle. To a given $x$ corresponds the counterclockwise rotation through the angle $\varphi$, where $|\varphi| = |x|$. This parametrization is almost unique: two different points of the sphere specify the same rotation only if these are $x$ and $-x$ with $|x| = \pi$ (since rotations through the angles $\pi$ and $-\pi$ provide the same motion). Therefore, SO(3) is obtained from the sphere $|x| \leq \pi$ by gluing every two opposite points of the boundary into a single point. Thus we have shown that $\mathrm{SO}(3) \approx \mathbb{R}P^3$ (i.e., these manifolds are diffeomorphic).

The groups $\mathrm{SO}(p, q)$ for nonzero $p$ and $q$ are not connected surfaces in $\mathrm{M}(p + q, \mathbb{R})$. For example, for $p = 1$ they contain orthochronous and nonortho-
chronous transformations, which cannot be joined by curves in $\mathrm{SO}(1, q)$.

**5.2.2. Complex surfaces and subgroups of** $\mathrm{GL}(n,\mathbb{C})$**.** Now we turn to the complex case. Denote by $\mathrm{M}(n,\mathbb{C})$ the space of all $n\times n$ matrices with elements in $\mathbb{C}$. The group $\mathrm{GL}(n,\mathbb{C})$ of invertible matrices is the domain in $\mathrm{M}(n,\mathbb{C})=\mathbb{C}^{n^2}$ specified by the condition $\det A\neq 0$.

A surface $M$ in $\mathbb{C}^n$ is said to be *complex* if, in a neighborhood of each of its points, it is determined by the equations

$$F_1(z^1,\dots,z^n)=\dots=F_{n-k}(z^1,\dots,z^n)=0 \tag{5.18}$$

with complex-analytic functions $F_1,\dots,F_{n-k}$. A point of the surface is called nonsingular if the rank of the matrix $\left(\frac{\partial F_j}{\partial z^l}\right)$ at this point is maximal and equal to $n-k$. The number $k$ is then called the complex dimension of $M$, $\dim_{\mathbb{C}} M=k$. By the implicit function theorem for complex-analytic functions, if a minor, say,

$$\left(\frac{\partial F_j}{\partial z^l}\right)_{j=1,\dots,n-k,\ l=k+1,\dots,n}$$

does not vanish at $x\in M$, then the surface can be locally specified as the graph of a mapping

$$z^{k+1}=f_1(z^1,\dots,z^k),\quad \dots,\quad z^n=f_{n-k}(z^1,\dots,z^k),$$

where $f_1,\dots,f_{n-k}$ are complex-analytic functions.

We may also consider the corresponding real space $\mathbb{R}^{2n}$, in which equations (5.18) specify a $(2k)$-dimensional surface $M$ with $\dim M=2\dim_{\mathbb{C}} M$. In this case a point $x\in M$ is nonsingular if and only if it is nonsingular as a point of the complex surface. The only distinctive feature of this surface $M^{2k}\subset\mathbb{R}^{2n}$ is that we can introduce complex coordinates on it and hence define complex-analytic functions.

An example of a complex surface in $\mathbb{C}^{n^2}$ is given by the group $\mathrm{SL}(n,\mathbb{C})$ formed by all matrices with determinant one. It is specified by the equation

$$\det A=1,$$

where the determinant is a polynomial, and hence, a complex-analytic function of matrix elements. This equation consists of two real equations, $\mathrm{Re}\det A=1$ and $\mathrm{Im}\det A=0$. The group $\mathrm{SL}(n,\mathbb{C})$ is of complex dimension $n^2-1$, and the tangent space to it at the identity consists of zero-trace matrices. These facts are proved along the same lines as their analogs for $\mathrm{SL}(n,\mathbb{R})$.

The group $\mathrm{U}(n) \subset \mathrm{M}(n, \mathbb{C})$ is described as the set of zeros of complex-valued, but not complex-analytic functions:

$$
(5.19) \qquad \begin{gathered}
f_{jk}(A) - \delta_{jk} = 0, \quad j, k = 1, \ldots, n, \quad \text{or} \quad A^\top \bar{A} - 1 = 0, \\
A = (a^i_j), \quad f_{jk}(A) = \sum_{l=1}^{n} a^j_l \bar{a}^k_l .
\end{gathered}
$$

Therefore, $\mathrm{U}(n)$ is not a complex surface. Each of the equations $f_{jk}(A) = \delta_{jk}$ determines two real equations:

$$
\operatorname{Re} f_{jk} = \delta_{jk}, \qquad \operatorname{Im} f_{jk} = 0.
$$

Since $f_{jk} = \bar{f}_{kj}$ and $\operatorname{Im} f_{jj} = 0$, we have $n^2 = n + 2\,\frac{n(n-1)}{2}$ different equations. Let $A(t)$ be a curve in $\mathrm{U}(n)$ such that $A(0) = 1$. Consider its Taylor expansion at $t = 0$:

$$
A(t) = 1 + Xt + O(t^2).
$$

From (5.19) we have

$$
A^\top \bar{A} = (1 + X^\top t + O(t^2))(1 + \bar{X}t + O(t^2)) = 1 + (X^\top + \bar{X})t + O(t^2) = 1.
$$

Therefore, equations (5.6) for the group $\mathrm{U}(n)$ specify the space of all skew-Hermitian matrices,

$$
x^j_k = -\bar{x}^k_j .
$$

The dimension of this space equals $n^2 = n + 2\,\frac{n(n-1)}{2}$. Now Lemmas 5.3 and 5.7 imply the following result.

**Theorem 5.20.** *The group* $\mathrm{U}(n)$ *is an* $n^2$*-dimensional nonsingular surface in* $\mathrm{M}(n, \mathbb{C}) = \mathbb{C}^{n^2}$. *The tangent space to* $\mathrm{U}(n)$ *at the identity consists of all skew-Hermitian matrices.*

It can be proved in a similar way that $\mathrm{SU}(n)$ is an $(n^2 - 1)$-dimensional nonsingular surface and its tangent space coincides with the space of skew-Hermitian zero-trace matrices.

EXAMPLE. The group $\mathrm{SU}(2)$ consists of matrices

$$
\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \quad |a|^2 + |b|^2 = 1.
$$

Let $a = x + iy$, $b = u + iv$. The equation $|a|^2 + |b|^2 = 1$ specifies a sphere of radius 1 in $\mathbb{R}^4$ with coordinates $x$, $y$, $u$, $v$. Recall that $\mathrm{SO}(3) = \mathrm{SU}(2)/\pm 1$ (Theorem 4.4).

### 5.2.3. Groups of affine transformations and the Heisenberg group.

The Lie subgroups of the group $\mathrm{GL}(n, \mathbb{R})$ may have a very simple structure. We will indicate here an important example, namely, the *Heisenberg group* consisting of all matrices of the form

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}, \qquad x, y, z \in \mathbb{R},$$

with usual multiplication operation. It is diffeomorphic to the Euclidean space $\mathbb{R}^3$, but the group operation is not commutative:

$$(x, y, z) \cdot (x', y', z') = (x + x', y + y', z + z' + xy').$$

Another example of a matrix group which is not a group of motions is provided by the group of affine transformations $\mathrm{A}(n)$. Let $x^1, \ldots, x^n$ be coordinates in the Cartesian space $\mathbb{R}^n$. Affine transformations have the form

$$(A, b)\colon x \to Ax + b, \quad A \in \mathrm{GL}(n), \quad b \in \mathbb{R}^n. \tag{5.20}$$

Consider the Cartesian space $\mathbb{R}^{n+1}$ of larger dimension with coordinates $x^1, \ldots, x^{n+1}$ and imbed $\mathbb{R}^n$ into it as a hyperplane $x^{n+1} = 1$:

$$(x^1, \ldots, x^n) \to (x^1, \ldots, x^n, 1).$$

To each affine transformation $(A, b)$ we assign the linear transformation $C$ specified by the $(n+1) \times (n+1)$ matrix:

$$(A, b) \to C = \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}. \tag{5.21}$$

The affine transformation (5.20) is obtained as the restriction of this transformation to the hyperplane $\mathbb{R}^n = \{x^{n+1} = 1\}$. It can be verified by a direct calculation that the mapping $\mathrm{A}(n) \to \mathrm{GL}(n+1)$ constructed in this way is a homomorphism of groups. Thus we have proved the following theorem.

**Theorem 5.21.** *The mapping* (5.21) *specifies a homomorphic imbedding of the group of affine transformations* $\mathrm{A}(n)$ *into the group* $\mathrm{GL}(n+1)$.

If the matrix $A$ in (5.20) is orthogonal, then the affine transformation preserves the distance between points defined by the formula

$$\rho(x, y) = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2},$$

where $x = (x^1, \ldots, x^n)$, $y = (y^1, \ldots, y^n) \in \mathbb{R}^n$. The subgroup $\mathrm{E}(n)$ formed by affine transformations (5.20) with orthogonal matrices $A$ is the group of motions of the Euclidean space.

Taking Theorem 5.16 into account we obtain the following result.

**Corollary 5.8.** *The affine group* $\mathrm{A}(n)$ *is an* $(n^2+n)$*-dimensional regular surface in* $\mathrm{M}(n+1,\mathbb{R}) = \mathbb{R}^{(n+1)^2}$*. The group of motions* $\mathrm{E}(n)$ *of the Euclidean space* $\mathbb{R}^n$ *is an* $\left(\frac{n(n-1)}{2}+n\right)$*-dimensional regular surface in* $\mathbb{R}^{(n+1)^2}$*.*

**5.2.4. Exponential mapping.** The tangent spaces to the group $\mathrm{GL}(n,\mathbb{C})$ and its subgroups enable one to recover the groups, at least locally, and introduce convenient local coordinates on them. This can be done with the aid of the exponential mapping,

$$\exp\colon T \to \mathrm{GL}(n,\mathbb{C}), \quad \exp(0) = 1,$$

where $T = \mathrm{M}(n,\mathbb{C})$ is the tangent space to $\mathrm{GL}(n,\mathbb{C})$ at the identity. This mapping is determined by the formal series

$$\exp X = 1 + \frac{X}{1!} + \frac{X^2}{2!} + \cdots \tag{5.22}$$

and is also denoted by $e^X = \exp X$.

**Lemma 5.9.** 1. *The series* (5.22) *converges for all* $X \in \mathrm{M}(n,\mathbb{C})$.

2. *If the matrices* $X$ *and* $Y$ *commute, i.e.,* $XY = YX$*, then*

$$\exp(X+Y) = \exp X \, \exp Y = \exp Y \, \exp X. \tag{5.23}$$

3. *For any matrix* $X \in \mathrm{M}(n,\mathbb{C})$,

$$\exp X \cdot \exp(-X) = 1, \tag{5.24}$$

$$\exp(X^\top) = (\exp X)^\top. \tag{5.25}$$

4. *For any* $X \in \mathrm{M}(n,\mathbb{C})$ *and* $A \in \mathrm{GL}(n,\mathbb{C})$,

$$\exp(AXA^{-1}) = A(\exp X)A^{-1}. \tag{5.26}$$

5. *For any* $X \in \mathrm{M}(n,\mathbb{C})$,

$$\det e^X = e^{\mathrm{Tr}\, X}. \tag{5.27}$$

**Proof.** Since

$$\left|\frac{X^k}{k!} + \frac{X^{k+1}}{(k+1)!} + \cdots + \frac{X^{k+l}}{(k+l)!}\right| \le \frac{|X|^k}{k!} + \cdots + \frac{|X|^{k+l}}{(k+l)!},$$

convergence of the series (5.22) follows from convergence of the series for $e^{|X|}$. If $XY = YX$, then

$$\exp X \exp Y = \left(\sum_{k=0}^{\infty} \frac{X^k}{k!}\right)\left(\sum_{l=0}^{\infty} \frac{Y^l}{l!}\right) = \sum_{m=0}^{\infty} \frac{1}{m!}\Big(\sum_{k+l=m} \frac{m!}{k!\,l!} X^k Y^l\Big)$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!}(X+Y)^m = \exp(X+Y).$$

Since $X$ and $-X$ commute, equalities (5.23) imply (5.24). The formulas (5.25) and (5.26) are obvious. To prove (5.27), we reduce the matrix $X$ to the Jordan form $\tilde{X}$:

$$\tilde{X} = AXA^{-1}.$$

Obviously, $\operatorname{Tr} \tilde{X} = \operatorname{Tr} X$ and (5.26) implies that

$$\det e^{\tilde{X}} = \det e^X.$$

The matrix $\tilde{X}$ is upper triangular with coefficients $\lambda_1, \dots, \lambda_n$ on the diagonal. We have

$$\det e^{\tilde{X}} = e^{\lambda_1} \cdots e^{\lambda_n} = e^{\lambda_1 + \dots + \lambda_n} = e^{\operatorname{Tr} \tilde{X}},$$

which implies (5.27). □

**Theorem 5.22.** 1. *If $X \in \mathrm{M}(n, \mathbb{R})$ (or $X \in \mathrm{M}(n, \mathbb{C})$) and $\operatorname{Tr} X = 0$, then $\exp X \in \mathrm{SL}(n, \mathbb{R})$ (or $X \in \mathrm{SL}(n, \mathbb{C})$, respectively).*

2. *If $X \in \mathrm{M}(n, \mathbb{R})$ is a skew-symmetric matrix, i.e., $X^\top = -X$, then $\exp X \in \mathrm{O}(n)$.*

3. *If $X \in \mathrm{M}(n, \mathbb{C})$ is a skew-Hermitian matrix, i.e., $\bar{X}^\top = -X$, then $\exp X \in \mathrm{U}(n)$.*

**Proof.** 1. Let $\operatorname{Tr} X = 0$ and $A(t) = \exp(Xt)$. Since $t_1 X$ and $t_2 X$ commute for all $t_1$, $t_2 \in \mathbb{R}$, we have

$$A(t_1 + t_2) = A(t_1) A(t_2).$$

Setting $f(t) = \det A(t)$, we obtain

$$f(t_1 + t_2) = f(t_1) f(t_2), \quad f(0) = 1.$$

This implies that

$$\frac{df(t)}{dt} = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} = f(t) \cdot \lim_{\Delta t \to 0} \frac{f(\Delta t) - 1}{\Delta t} = Cf(t),$$

where the constant $C$ equals

$$\begin{aligned} C = \lim_{\Delta t \to 0} \frac{f(\Delta t) - 1}{\Delta t} &= \lim_{\Delta t \to 0} \frac{\det(1 + X\Delta t + O(\Delta t^2)) - 1}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\operatorname{Tr} X \, \Delta t + O(\Delta t^2)}{\Delta t} = \operatorname{Tr} X = 0. \end{aligned}$$

We obtain that $\frac{df}{dt} = 0$ everywhere and $f(0) = 1$. Therefore, $\det A(t) \equiv 1$.

2. Since $X$ is skew-symmetric, i.e., $X^\top = -X$, the matrices $X$ and $X^\top$ commute. Hence

$$A^\top A = \exp X^\top \exp X = \exp(X^\top + X) = 1, \quad A \in \mathrm{O}(n).$$

3. Since $X$ is skew-Hermitian, $X^\top = -\bar{X}$, we have

$$A^\top \bar{A} = \exp X^\top \exp \bar{X} = \exp(X^\top + \bar{X}) = 1, \quad A \in \mathrm{U}(n).$$

The proof is completed. □

**Lemma 5.10.** *If* $A = 1 + Y \in \mathrm{GL}(n, \mathbb{C})$ *and* $|Y| < 1$, *then*

$$A = \exp X,$$
$$X = \log A = Y - \frac{Y^2}{2} + \frac{Y^3}{3} + \cdots + (-1)^{n-1}\frac{Y^n}{n} + \cdots .$$

**Proof.** The series for $\log A$ converges if the series

$$|Y| + \frac{|Y|^2}{2} + \cdots + \frac{|Y|^n}{n} + \cdots ,$$

which is the Taylor series for $\log(1 - z)$, $z = |Y|$, is convergent. It is well known that the Taylor series for $\log(1 - z)$ converges for $|z| < 1$. Therefore, for $|Y| < 1$, the series for $X = \log A$ converges. Substituting it into $\exp X$ we obtain $A = \exp X$. □

**Corollary 5.9.** *In a neighborhood of the unit element of* $\mathrm{GL}(n, \mathbb{R})$ *the mapping* exp *is one-to-one and the entries* $x^i_j = a^i_j - \delta^i_j$ *determine local coordinates.*

A smooth curve $r(t)$, $t \in \mathbb{R}$, in the group $\mathrm{GL}(n, \mathbb{C})$ (or in its subgroup) is called a *one-parameter subgroup* of the group $\mathrm{GL}(n, \mathbb{C})$ if

$$r(t_1 + t_2) = r(t_1)r(t_2)$$

for all $t_1, t_2 \in \mathbb{R}$. Obviously, $r(0) = 1 \in \mathrm{GL}(n, \mathbb{C})$.

**Lemma 5.11.** *If* $r(t)$ *is a one-parameter subgroup of the group* $\mathrm{GL}(n, \mathbb{C})$, *then* $r(t) = e^{Xt}$ *for some matrix* $X \in \mathrm{M}(n, \mathbb{C})$.

**Proof.** We have

$$\frac{dr(t)}{dt} = \lim_{\Delta t \to 0} \frac{r(t + \Delta t) - r(t)}{\Delta t} = r(t)\Big(\lim_{\Delta t \to 0} \frac{r(\Delta t) - 1}{\Delta t}\Big) = r(t)X,$$

where

$$X = \lim_{\Delta t \to 0} \frac{r(\Delta t) - 1}{\Delta t}.$$

But

$$\frac{de^{Xt}}{dt} = e^{Xt}X.$$

Therefore, the functions $r(t)$ and $e^{Xt}$ satisfy the same ordinary differential equation

$$\frac{df}{dt} = fX$$

in $\mathbb{R}^{2n^2} = \mathbb{C}^{n^2} = \mathrm{M}(n, \mathbb{C})$ with the same initial condition

$$f(0) = 1.$$

By the theorem on the existence and uniqueness of solutions of ordinary differential equations we obtain

$$r(t) = e^{Xt}$$

for all $t \in \mathbb{R}$. □

**Lemma 5.12.** *Let $F\colon G \to H$ be a smooth homomorphism of groups that are nonsingular surfaces in* $\mathrm{M}(n, \mathbb{R})$,

$$F(x^{-1}) = [F(x)]^{-1}, \qquad F(xy) = F(x)F(y), \quad x, y \in G.$$

*Then $F$ maps one-parameter subgroups into one-parameter subgroups,*

$$F(e^{Xt}) = e^{Yt},$$

*where $Y = F_*X$ and $F_*$ is the differential of the mapping $F$ at the identity.*

**Proof.** Obviously, $r(t) = F(e^{Xt})$ is a one-parameter subgroup. Hence $r(t) = e^{Yt}$. The tangent vector to the curve $r(t)$ at $t = 0$ is

$$Y = \frac{dr}{dt}(0) = F_*\Big(\frac{de^{Xt}}{dt}\Big|_{t=0}\Big) = F_*X.$$

□

For some (but not all) subgroups in $\mathrm{GL}(n, \mathbb{R})$, the image of the exponential mapping covers the entire group. We will deduce this property for $\mathrm{SO}(n)$ from Lemma 5.8.

**Lemma 5.13.** *Any element of the group* $\mathrm{SO}(n)$ *is representable as* $\exp X$ *for some skew-symmetric matrix $X$.*

**Proof.** Let $A \in \mathrm{SO}(n)$. By Lemma 5.8, we can find in $V$ an orthonormal basis in which the transformation $A\colon V \to V$ is specified by a matrix $\tilde{A}$ of the form (5.17). We have

$$\tilde{A} = BAB^{-1}, \quad B \in \mathrm{SO}(n).$$

One can easily calculate that

$$\exp\left[\varphi\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right] = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}.$$

Hence

$$\tilde{A} = \exp\tilde{X} = \exp\begin{pmatrix} 0 & 0 & & 0 \\ 0 & \varphi_1 J & & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & & \varphi_k J \end{pmatrix}, \quad \text{where } J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The matrix $\tilde{X}$ is skew-symmetric. Let us show that $X = B^{-1}\tilde{X}B$ is also skew-symmetric. Indeed, since $B^\top = B^{-1}$, we have

$$(B^{-1}\tilde{X}B)^\top = B^\top \tilde{X}^\top (B^{-1})^\top = B^{-1}\tilde{X}^\top B = -B^{-1}\tilde{X}B.$$

It remains to note that

$$A = B^{-1}(\exp \tilde{X})B = \exp(B^{-1}\tilde{X}B) = \exp X.$$

$\square$

## 5.3. Quaternions and groups of motions

**5.3.1. Algebra of quaternions.** Let $\mathbb{H}$ be the four-dimensional real vector space consisting of vectors

$$q = a + bi + cj + dk,$$

where $a, b, c, d \in \mathbb{R}$ are the coordinates of the vectors. Define an associative bilinear multiplication on this space by the following rules:

1) $(\lambda q_1)(\mu q_2) = (\lambda\mu) q_1 q_2$, $\lambda, \mu \in \mathbb{R}$, $q_1, q_2 \in \mathbb{H}$;
2) $(q_1 q_2) q_3 = q_1 (q_2 q_3)$ (associativity);
3) for the basis vectors,

$$ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j,$$
$$i^2 = j^2 = k^2 = -1.$$

The space $\mathbb{H}$ with multiplication so defined is called the *algebra of quaternions*. It can be realized in $\mathrm{M}(2, \mathbb{C})$ by setting

$$q = a + bi + cj + dk \to A(q) = \begin{pmatrix} a - id & -bi - c \\ -bi + c & a + id \end{pmatrix}.$$

One can easily verify that

$$A(q_1 q_2) = A(q_1)A(q_2), \qquad A(q_1 + q_2) = A(q_1) + A(q_2).$$

The matrices

$$\sigma_0 = A(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 = iA(i) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$\sigma_2 = iA(j) = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = iA(k) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are called the *Pauli matrices*. They satisfy the equalities

$$\sigma_0 = 1, \qquad \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1,$$
$$\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3, \qquad \sigma_2\sigma_3 = -\sigma_3\sigma_2 = i\sigma_1, \qquad \sigma_3\sigma_1 = -\sigma_1\sigma_3 = i\sigma_2.$$

On the algebra of quaternions, we define the conjugation operation

$$q = a + bi + cj + dk \to \bar{q} = a - bi - cj - dk$$

and the corresponding norm

$$|q|^2 = q\bar{q} = a^2 + b^2 + c^2 + d^2.$$

Since

$$A(\bar{q}) = \overline{A(q)}^{\top},$$

we have

$$\overline{q_1 + q_2} = \bar{q}_1 + \bar{q}_2, \qquad \overline{q_1 \cdot q_2} = \bar{q}_2 \cdot \bar{q}_1.$$

Note also that

$$|q|^2 = \det A(q).$$

Therefore

$$|q_1 q_2| = |q_1||q_2|. \tag{5.28}$$

Since this norm is generated by the scalar product

$$\langle q_1, q_2 \rangle = a_1 a_2 + b_1 b_2 + c_1 c_2 + d_1 d_2,$$

it has the usual properties

$$|q_1 + q_2| \leq |q_1| + |q_2|, \qquad |\lambda q| = |\lambda||q| \quad \text{for } \lambda \in \mathbb{R}, \qquad |q| \geq 0 \quad \text{for } q \neq 0.$$

Therefore, $\mathbb{H}$ is a *normed algebra* (the formal definitions of abstract and normed algebras will be given in 6.1.10). In $\mathbb{H}$, the division operation is also defined: to each nonzero element $q \in \mathbb{H}$ corresponds a unique inverse,

$$q^{-1} = \frac{\bar{q}}{|q|^2}, \qquad qq^{-1} = q^{-1}q = 1.$$

A finite-dimensional vector space with bilinear multiplication such that the equality $xy = 0$ implies that at least one of the factors is equal to zero is called a *division algebra.* This condition (that there are no divisors of zero) implies that for each $a \neq 0$, the mappings $x \to ax$ and $x \to xa$ are invertible linear transformations. Real division algebras exist only in dimensions $1, 2, 4$, and $8$ (which is a complicated topological theorem).

If a division algebra contains a unity (an element multiplication by which is the identity transformation) and the multiplication is associative, then it can be proved that there are only three such algebras over the field of real numbers $\mathbb{R}$, namely, the field of real numbers $\mathbb{R}$ itself, the field of complex numbers $\mathbb{C}$, and the algebra of quaternions $\mathbb{H}$.

**5.3.2. The groups SO(3) and SO(4).** Let $x = a - id$, $y = -c - ib$. Then

$$A(q) = \begin{pmatrix} x & y \\ -\bar{y} & \bar{x} \end{pmatrix}, \qquad |q|^2 = |x|^2 + |y|^2.$$

The equality (5.28) implies that the quaternions for which $|q| = 1$ form the multiplicative group of unit quaternions $\mathbb{H}_1$. It is easily seen that this group is isomorphic to SU(2).

Denote by $\mathbb{H}_0$ the space of imaginary quaternions:

$$\bar{q} = -q, \qquad q = xi + yj + zk.$$

The metric in $\mathbb{H}_0$ is defined by the formula

$$|q|^2 = -q^2 = x^2 + y^2 + z^2.$$

Therefore, $\mathbb{H}_0$ can be identified with the three-dimensional Euclidean vector space.

**Lemma 5.14.** *If $|q| = 1$, then the transformation*

$$\alpha_q : x \to qx\bar{q}, \quad x \in \mathbb{H}_0 = \mathbb{R}^3, \quad q \in \mathbb{H}_1 = \mathrm{SU}(2),$$

*is a rotation of the Euclidean space $\mathbb{R}^3$, and the quaternions $\pm q$ determine the same rotation.*

**Proof.** We have $\bar{x} = -x$, hence

$$\overline{\alpha_q(x)} = \overline{qx\bar{q}} = q\bar{x}\bar{q} = -qx\bar{q}.$$

Therefore, $qx\bar{q} \in \mathbb{H}_0$. The mapping $\alpha_q$ is linear and $\alpha_q(0) = 0$. Moreover, we have

$$|qx\bar{q}| = |q||x||\bar{q}| = |x|,$$

hence $\alpha_q$ is a motion and is determined by a matrix $\alpha_q \in \mathrm{O}(3)$. The function $f(q) = \det \alpha_q$ is a continuous function of $q$ and takes only two values $\pm 1$. But $f(1) = 1$ and $\mathrm{SU}(2) \approx S^3$ is a connected surface. Therefore, $\det \alpha_q \equiv 1$ and $\alpha_q \in \mathrm{SO}(3)$ for $q \in \mathbb{H}_1$. □

The group $\mathrm{SU}(2) \times \mathrm{SU}(2)$ consists of all pairs $(g_1, g_2)$, where $g_1, g_2 \in \mathrm{SU}(2)$, with multiplication operation

$$(g_1, g_2) \cdot (h_1, h_2) = (g_1 h_1, g_2 h_2).$$

Obviously, this group is isomorphic to $\mathbb{H}_1 \times \mathbb{H}_1$.

The space $\mathbb{H}$ with the norm $|q|$ is identified in a natural way with the space of radius-vectors of points of the Euclidean space $\mathbb{R}^4$.

**Lemma 5.15.** *If $|p| = |q| = 1$, then the transformation*

$$\alpha_{p,q} : x \to px\bar{q}, \quad x \in \mathbb{H} = \mathbb{R}^4, \quad p, q \in \mathbb{H}_1 = \mathrm{SU}(2),$$

*is a rotation of the Euclidean space $\mathbb{R}^4$.*

The proof of this lemma is similar to that of Lemma 5.14.

Thus we have constructed smooth mappings

$$q \in \mathrm{SU}(2) \to \alpha_q \in \mathrm{SO}(3), \tag{5.29}$$

$$(p, q) \in \mathrm{SU}(2) \times \mathrm{SU}(2) \to \alpha_{p,q} \in \mathrm{SO}(4), \tag{5.30}$$

which are at the same time homomorphisms of groups.

It can be proved by simple calculations that these mappings have maximal rank at the identity elements of the groups, and hence determine isomorphisms of tangent spaces at the identities.

Therefore, any vector $Y$ tangent to SO(3) or SO(4) at the identity has the form $F_*X$, where $F$ is a mapping given by (5.29) or (5.30), and $X$ is a tangent vector to SU(2) or SU(2) × SU(2) at the identity. The image of the one-parameter group $e^{Xt}$ is the group $e^{Yt}$. Since the exponential mapping covers the entire SO($n$), the images of the mappings (5.29) and (5.30) cover the entire groups SO(3) and SO(4).

It is easily seen that

1) $\alpha_p = 1$ if and only if $p = \pm 1 \in \mathbb{H}_1 = \mathrm{SU}(2)$;

2) $\alpha_{p,q} = 1$ if and only if $p = q = \pm 1 \in \mathbb{H}_1 \times \mathbb{H}_1 = \mathrm{SU}(2) \times \mathrm{SU}(2)$.

Thus we have proved the following theorem.

**Theorem 5.23.** 1. *The mapping* (5.29) *is a smooth homomorphism of the group* SU(2) *onto* SO(3) *with kernel* $\pm 1$. *Therefore,*

$$\mathrm{SO}(3) \simeq \mathrm{SU}(2)/\pm 1.$$

2. *The mapping* (5.30) *is a smooth homomorphism of the group* SU(2) × SU(2) *onto* SO(4) *with kernel* $\pm 1 = (\pm 1, \pm 1)$. *Therefore,*

$$\mathrm{SO}(4) \simeq (\mathrm{SU}(2) \times \mathrm{SU}(2))/\pm 1,$$

$$\mathrm{SO}(4)/\pm 1 \simeq \mathrm{SO}(3) \times \mathrm{SO}(3).$$

The group SU(2) is diffeomorphic (as a smooth manifold) to the three-dimensional sphere. Since the projective spaces $\mathbb{R}P^n$ are quotient manifolds of spheres by $\mathbb{Z}_2$-actions, we obtain the following result.

**Corollary 5.10.** *The group* SO(3) *is diffeomorphic to the three-dimensional real projective space*:

$$\mathrm{SO}(3) \approx \mathbb{R}P^3.$$

**5.3.3. Quaternion-linear transformations.** Quaternions are closely related also to the groups Sp($n$) (not to be confused with symplectic groups Sp($n, \mathbb{R}$) defined in Section 2.1.2).

Let $\mathbb{H}^n$ be the $n$-dimensional quaternion space with basis $e_1, \dots, e_n$ and coordinates $(q_1, \dots, q_n)$ taking values in $\mathbb{H}$. Since a quaternion

$$q = a + bi + cj + dk$$

can be represented as

$$q = x + yj = x + j\bar{y}, \quad \text{where} \quad x = a + bi, \quad y = c + di,$$

$\mathbb{H}^n$ may be regarded as the space $\mathbb{C}^{2n}$ with basis $e_1, \dots, e_n, je_1, \dots, je_n$ and complex coordinates $x^1, \dots, x^n, y^1, \dots, y^n$.

A quaternion-linear transformation is an invertible transformation $\Lambda\colon \mathbb{H}^n \to \mathbb{H}^n$ specified by a matrix $\Lambda = (\lambda_l^k)$ by the formula

$$q^k \to q^l \lambda_l^k = \tilde{q}^k$$

(since the coordinates $q^k$ and the entries $\lambda_l^k$ belong to $\mathbb{H}$ and multiplication in $\mathbb{H}$ is noncommutative, the order of the factors is essential). This transformation is linear in the following sense:

$$\Lambda(\xi_1 + \xi_2) = \Lambda\xi_1 + \Lambda\xi_2, \qquad \Lambda(q\xi) = q(\Lambda\xi) \quad \text{for } q \in \mathbb{H}.$$

Note that in general

$$\Lambda(\xi q) \neq q(\Lambda\xi) \quad \text{and} \quad \Lambda(\xi q) \neq (\Lambda\xi)q.$$

Note also that a matrix $\Lambda = j \cdot 1$ corresponds to right multiplication of the coordinates of a point $\xi$ by $j$. Identifying $\mathbb{H}^n$ with $\mathbb{C}^{2n}$ we obtain the complex form of the transformation $\Lambda$:

$$x^k \to x^k a_l^k - y^l \bar{b}_l^k, \qquad y^k \to x^l b_l^k + y^l \bar{a}_l^k, \qquad \lambda_l^k = a_l^k + b_l^k j. \tag{5.31}$$

Quaternion-linear transformations form a group $\mathrm{GL}(n, \mathbb{H})$, and formulas (5.31) specify the complexification homomorphism

$$c\colon \mathrm{GL}(n, \mathbb{H}) \to \mathrm{GL}(2n, \mathbb{C}) \tag{5.32}$$

with

$$c(\Lambda) = \begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix}, \qquad \begin{pmatrix} x \\ y \end{pmatrix} \to c(\Lambda) \begin{pmatrix} x \\ y \end{pmatrix}, \quad \Lambda = A + Bj.$$

As in the case of the realization homomorphism $r$, the image of this homomorphism in $\mathrm{GL}(2n, \mathbb{C})$ is specified by the condition of commuting with (right) multiplication by the imaginary unit $j$:

$$\Lambda c(j) = c(j)\bar{\Lambda}, \quad c(j) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \Lambda \in c(\mathrm{Sp}(n)) \subset \mathrm{GL}(2n, \mathbb{C}).$$

Define the Hermitian scalar product on quaternion-valued vectors in $\mathbb{H}^n$,

$$\langle \xi_1, \xi_2 \rangle_{\mathbb{H}} = \sum_{k=1}^{n} \xi_1^k \bar{\xi}_2^k, \quad \xi_1 = (\xi_1^k), \quad \xi_2 = (\xi_2^k).$$

The group $\mathrm{Sp}(n)$ consists of all quaternion-linear transformations preserving this form,

$$\langle \Lambda\xi_1, \Lambda\xi_2 \rangle_{\mathbb{H}} = \langle \xi_1, \xi_2 \rangle_{\mathbb{H}} \quad \text{for} \quad \Lambda \in \mathrm{Sp}(n).$$

In complex coordinates $x^k$, $y^k$, where $\xi^k = x^k + y^k j$, $k = 1, \ldots, n$, the form $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ is written as

$$\langle \xi_1, \xi_2 \rangle_{\mathbb{H}} = \sum_{k=1}^{n} (x_1^k \bar{x}_2^k + y_1^k \bar{y}_2^k) + \sum_{k=1}^{n} (y_1^k x_2^k - x_1^k y_2^k) j.$$

Therefore, a transformation $\Lambda$ belongs to $\mathrm{Sp}(n)$ if it preserves both the Hermitian form $\langle \xi_1, \xi_2 \rangle_{\mathbb{C}} = \sum_{k=1}^{n} (x_1^k \bar{x}_2^k + y_1^k \bar{y}_2^k)$ and the symplectic form $\sum_{k=1}^{n} (y_1^k x_2^k - x_1^k y_2^k)$. Hence we conclude that

$$c(\mathrm{Sp}(n)) \subset \mathrm{U}(2n).$$

It is easily seen that the mapping (5.32) establishes the isomorphism

$$\mathrm{Sp}(1) = \mathrm{SU}(2).$$

## Exercises to Chapter 5

1. Show that there is no atlas on the sphere consisting of a single chart.

2. Construct a smooth one-to-one mapping of two smooth manifolds which is not a diffeomorphism.

3. Show that the set of all straight lines on the plane forms a smooth manifold which is homeomorphic to the Möbius strip.

4. Construct the imbeddings:

a) of the torus $T^n = S^1 \times \cdots \times S^1$ into $\mathbb{R}^{n+1}$;

b) of the product of spheres $S^k$ and $S^n$ into $\mathbb{R}^{k+n+1}$.

5. Show that a pair of intersecting straight lines on the plane is not a manifold.

6. Show that any smooth manifold has an atlas with each chart homeomorphic to a Euclidean space.

7. If the Möbius strip is cut along its middle line, will the resulting manifold be orientable? connected? What will happen if this procedure is repeated several times?

8. Construct an atlas of four charts on the torus $T^2$.

*Chapter 6*

# Groups of Motions

## 6.1. Lie groups and algebras

**6.1.1. Lie groups.** A smooth manifold $G$ is called a *Lie group* if it is a group such that the group operations, i.e., multiplication

$$G \times G \to G\colon (g,h) \to gh \in G$$

and inversion

$$G \to G\colon g \to g^{-1},$$

are smooth mappings.

A *homomorphism of Lie groups* $G$ and $H$ is a smooth mapping $\varphi\colon G \to H$ that at the same time is a homomorphism of groups. Two Lie groups are said to be isomorphic if there exists a homomorphism of these groups that at the same time is a diffeomorphism.

A simple example of a Lie group is provided by a vector space $\mathbb{R}^n$ with the usual addition operation

$$(x^1,\dots,x^n)\cdot(y^1,\dots,y^n) = (x^1+y^1,\dots,x^n+y^n).$$

If $\Lambda$ is a subgroup in $\mathbb{R}^n$ isomorphic to $\mathbb{Z}^n$, then the quotient space, which is the $n$-dimensional torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$, is a compact Lie group with respect to the same operation. The groups $\mathbb{R}^n$ and $T^n$ are commutative.

If a Lie group $H$ is imbedded into a Lie group $G$ and this imbedding of smooth manifolds is at the same time a homomorphism of groups, then $H$ is said to be a *Lie subgroup of the Lie group $G$*.

Lie subgroups of the Lie groups $\mathrm{GL}(n,\mathbb{R})$ are called *matrix Lie groups.* These are, e.g., the groups $\mathrm{GL}(n,\mathbb{R})$, $\mathrm{SL}(n,\mathbb{R})$ and the groups of motions $\mathrm{O}(n)$ and $\mathrm{U}(n)$.

The following example shows that Lie groups are not exhausted by matrix groups.

Consider the group of the following transformations of the real line:

$$x \to x + 2\pi t + i \log \frac{1 - \bar{z}e^{ix}}{1 - ze^{-ix}}, \tag{6.1}$$

where $t \in \mathbb{R}$, $z \in \mathbb{C}$ with $|z| < 1$, and log is the continuous branch of the logarithmic function for which $\log 1 = 0$. This group, to be denoted by $\widetilde{\mathrm{SL}}(2, \mathbb{R})$, is a three-dimensional Lie group with coordinates $t, \operatorname{Re} z, \operatorname{Im} z$.

**Theorem 6.1.** *The group $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ cannot be imbedded into any group $\mathrm{GL}(n, \mathbb{R})$, and so it is not a matrix Lie group.*

**Proof.** Suppose that the group $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ is a matrix Lie group and is imbedded into $\mathrm{GL}(n, \mathbb{R}) \subset \mathrm{GL}(n, \mathbb{C})$. The equation $z = 0$ determines a one-parameter group with parameter $t$, which is isomorphic to $\mathbb{R}$. By Lemma 5.11 any such group has the form $e^{Xt}$, where $X \in \mathrm{M}(n, \mathbb{C})$. Recall that for a group $G$, the subgroup $H$ of elements that commute with all elements of $G$ is called the center of $G$. One can easily verify that the center $Z$ of the group $\widetilde{\mathrm{SL}}(2, \mathbb{R})$ consists of the elements such that $z = 0$ and $t \in \mathbb{Z}$. Therefore, this group must admit imbedding into the one-parameter subgroup $e^{Xt}$.

Take a basis in $\mathbb{C}^n$ in which the matrix $X$ has the Jordan form. Combine all the Jordan boxes with the same eigenvalues into blocks $X_1, \dots, X_k$, where each block has the form

$$X_i = \begin{pmatrix} \lambda_i & a_1 & 0 & & 0 \\ 0 & \lambda_i & a_2 & & 0 \\ 0 & 0 & \dots & & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & & a_l \\ 0 & 0 & 0 & & \lambda_i \end{pmatrix}$$

with $a_1, \dots, a_l$ equal to 0 or 1. Then

$$e^{X_i t} = e^{\lambda_i t} \begin{pmatrix} 1 & a_1 t & \frac{1}{2} a_1 a_2 t^2 & & \frac{1}{l!} a_1 \cdots a_l t^l \\ 0 & 1 & a_2 t & & \frac{1}{(l-1)!} a_2 \cdots a_l t^{l-1} \\ 0 & 0 & \dots & & \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & & a_l t \\ 0 & 0 & 0 & & 1 \end{pmatrix}.$$

Select an element $A \in \widetilde{\mathrm{SL}}(2, \mathbb{R}) \subset \mathrm{GL}(n, \mathbb{C})$ that does not commute with at least one element of the group $e^{Xt}$. Since for $t = 1$ the matrices $e^X$ and $A$ commute, i.e., $e^X A = Ae^X$, the subspaces corresponding to the blocks $X_i$ are invariant relative to $A$. Indeed, the vectors $\xi$ of such a subspace are

characterized by the condition that $(e^X - e^{\lambda_i})^m \xi = 0$ for some $m$, but since $e^X$ and $A$ commute, we obtain $(e^X - e^{\lambda_i})^m A\xi = 0$. Therefore, the matrices $e^{Xt}$ and $A$ reduce to the same block-diagonal form

$$e^{Xt} = \begin{pmatrix} e^{X_1 t} & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & e^{X_k t} \end{pmatrix}, \qquad A = \begin{pmatrix} A_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & A_k \end{pmatrix}.$$

The conditions for them to commute are

$$[e^{X_1 t}, A_1] = e^{X_1 t} A_1 - A_1 e^{X_1 t} = 0, \ \dots, \ [e^{X_k t}, A_k] = e^{X_k t} A_k - A_k e^{X_k t} = 0,$$

and it is seen from the explicit form of these matrices that this is a system of polynomial equations for $t$. The system is nontrivial because the matrices $A$ and $e^{Xt}$ do not commute for some $t$ by the choice of $A$. Therefore, the system must have finitely many solutions. But this contradicts the fact that it is satisfied for all integers $t$. This contradiction proves that the group $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ cannot be imbedded into any group $\mathrm{GL}(n,\mathbb{R})$ and hence is not a matrix group. □

Consider the center $Z$ of the group $\widetilde{\mathrm{SL}}(2,\mathbb{R})$. It acts on the Lie group $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ by right translations $g \to gh$, and this action is discrete. Consider the quotient manifold $\widetilde{\mathrm{SL}}(2,\mathbb{R})/Z$.

In order to describe this quotient manifold, rewrite the action (6.1) in the form

$$w \to e^{2\pi i t} \frac{w - z}{1 - \bar{z} w}, \quad w = e^{ix}.$$

Thus we obtain a homomorphism of the group $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ onto the group of all linear-fractional transformations of the disk $|w| < 1$. It was shown in Section 4.3.3 that the latter group is isomorphic to $\mathrm{SL}(2,\mathbb{R})/\pm 1$. The kernel of this homomorphism is the subgroup $Z$ that corresponds to the transformations such that $z = 0$, $t \in \mathbb{Z}$, and is isomorphic to $\mathbb{Z}$; a homomorphism is a smooth mapping. Thus we have proved the following result.

**Theorem 6.2.** *The quotient manifolds $\widetilde{\mathrm{SL}}(2,\mathbb{R})/Z$ and $\mathrm{SL}(2,\mathbb{R})/\pm 1$ are diffeomorphic, and this diffeomorphism is determined by the smooth homomorphism $\widetilde{\mathrm{SL}}(2,\mathbb{R}) \to \mathrm{SL}(2,\mathbb{R})/\pm 1$. The kernel of this homomorphism is the subgroup $Z$ , which is the center of the group $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ and is isomorphic to $\mathbb{Z}$.*

**6.1.2. Lie algebras.** Let $G$ be a Lie group, and let $x^1, \dots, x^n$ be local coordinates in a neighborhood of the identity element 1 of the group $G$ with $1 = (0, \dots, 0)$. For brevity we denote by $V = T_1 G$ the tangent space at the identity.

In the coordinates $(x^i)$, multiplication is written in terms of smooth functions $\varphi^i$, $i = 1, \dots, n$, as follows:

$$x \cdot y = (\varphi^1(x,y), \dots, \varphi^n(x,y)),$$

where $x = (x^1, \dots, x^n)$, $y = (y^1, \dots, y^n)$, and

$$\varphi^i(x,y) = x^i + y^i + b^i_{jk} x^j y^k + (\text{terms of order} \geq 3).$$

Let $\xi$ and $\eta$ be tangent vectors at the identity of the group, and let $\xi = (\xi^1, \dots, \xi^n)$, $\eta = (\eta^1, \dots, \eta^n)$ be their coordinates in the coordinate system $(x^i)$. Define the commutator $[\xi, \eta] \in V$ of these vectors by the formula

$$[\xi, \eta]^i = (b^i_{jk} - b^i_{kj}) \xi^j \eta^k. \tag{6.2}$$

**Lemma 6.1.** *The commutator*

$$[\cdot,\cdot] \colon V \times V \to V, \quad (\xi, \eta) \to [\xi, \eta],$$

*is a linear mapping in both arguments, is skew-symmetric, i.e.,*

$$[\xi, \eta] = -[\eta, \xi],$$

*and satisfies the Jacobi identity*

$$[\xi, [\eta, \chi]] + [\eta, [\chi, \xi]] + [\chi, [\xi, \eta]] = 0. \tag{6.3}$$

**Proof.** We must prove only the Jacobi identity because the rest is obvious from the definition of the commutator. The group multiplication is associative,

$$(xy)z = x(yz).$$

Let us write down the Taylor expansions for all components of the left- and right-hand sides of this equality about the point $x = y = z = 1$. The coefficients of $x^j y^k z^l$ in these expansions for the $i$th component, $i = 1, \dots, n$, coincide,

$$b^i_{ml} b^m_{jk} = b^i_{jm} b^m_{kl}.$$

Using this equality, the Jacobi identity can be verified by direct substitution of (6.2) into (6.3). □

Consider a simple example. Let $G = \mathrm{GL}(n, \mathbb{R})$. Each matrix in $\mathrm{GL}(n, \mathbb{R})$ is representable as $1 + A$, where the entries of $A$ may be taken for local coordinates in a neighborhood of the identity. The Taylor formula for multiplication reduces to a quadratic polynomial,

$$(1 + A)(1 + B) = 1 + A + B + AB.$$

Therefore, the commutator in the tangent space at the identity to the group $\mathrm{GL}(n, \mathbb{R})$ has a very simple form:

$$[A, B] = AB - BA, \tag{6.4}$$

where $A, B$ are matrices of the $n$th order. In this case the Jacobi identity can be verified very easily.

**Lemma 6.2.** *The commutation operation on the space* $\mathrm{M}(n, \mathbb{R})$ *of all matrices of order* $n$ *satisfies the Jacobi identity,*

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0. \tag{6.5}$$

**Proof.** Let us write down the terms in the left-hand side of (6.5) in detail:

$$[A, [B, C]] = A[B, C] - [B, C]A = ABC - ACB - BCA + CBA,$$
$$[C, [A, B]] = CAB - CBA - ABC + BAC,$$
$$[B, [C, A]] = BCA - BAC - CAB + ACB.$$

These expressions sum up to zero. □

Now we come to the following important notion.

A vector space $V$ with a binary operation

$$V \times V \to V, \quad (\xi, \eta) \to [\xi, \eta], \tag{6.6}$$

is called a *Lie algebra* if this operation is bilinear (i.e., linear in each variable), skew-symmetric, i.e.,

$$[\xi, \eta] = -[\eta, \xi],$$

and satisfies the *Jacobi identity*

$$[\xi, [\eta, \zeta]] + [\eta, [\zeta, \xi]] + [\zeta, [\xi, \eta]] = 0. \tag{6.7}$$

The operation (6.6) is referred to as the *commutator* in the Lie algebra.

**Corollary 6.1.** *The tangent space of a Lie group at the identity is a Lie algebra with respect to the operation* (6.2).

*In particular, the matrix algebra* $\mathrm{M}(n, \mathbb{R})$ *is a Lie algebra (over the field* $\mathbb{R}$ *of real numbers) with respect to the operation of matrix commutation* (6.4).

The tangent space at the identity of a Lie group $G$ with operation (6.2) is called the *Lie algebra of the Lie group.*

For a Lie algebra $V$ there is a mapping which associates with each element of it a linear operator on the algebra $V$. Namely, let $\xi \in V$. Define the linear mapping

$$\operatorname{ad} \xi \colon V \to V$$

by the formula

$$\operatorname{ad} \xi(\eta) = [\xi, \eta].$$

A linear operator $A \colon V \to V$ on a Lie algebra is referred to as a *differentiation* of the algebra $V$ if it satisfies the *Leibniz identity*

$$A([\xi, \eta]) = [A(\xi), \eta] + [\xi, A(\eta)].$$

The Jacobi identity (6.7) is equivalent to the following assertion: for any element $\xi \in V$, the mapping $\operatorname{ad}\xi$ is a differentiation of the Lie algebra $V$.

If $e_1, \dots, e_n$ is a basis in a finite-dimensional Lie algebra, then the multiplication in this basis is specified by the *structural constants* $c_{ij}^k$:

$$[e_i, e_j] = c_{ij}^k e_k$$

(as usual, the rule of summation over repeated indices is used). The condition that the commutator is skew-symmetric implies that the structural constants $c_{ij}^k$ are skew-symmetric with regard to subscripts:

$$c_{ij}^k = -c_{ji}^k \quad \text{for all } i, j, k. \tag{6.8}$$

The Jacobi identity (6.7) can be rewritten as a system of quadratic equations for structural constants. Indeed, it suffices to check this assertion for the basis vectors. In this case the equation

$$[e_i, [e_j, e_k]] + [e_j, [e_k, e_i]] + [e_k, [e_i, e_j]] = 0$$

rewrites as the system of equations

$$c_{il}^m c_{jk}^l + c_{jl}^m c_{ki}^l + c_{kl}^m c_{ij}^l = 0 \tag{6.9}$$

for $m = 1, \dots, \dim V$.

The simplest solution to equations (6.8) and (6.9) is trivial: $c_{ij}^k = 0$. It corresponds to the trivial commutator $[\xi, \eta] = 0$ for all $\xi, \eta \in V$. A Lie algebra with such a commutator is said to be *commutative*.

As we pointed out in Section 1.4.3, the three-dimensional Euclidean space $\mathbb{R}^3$ is a Lie algebra relative to the vector product operation. This algebra is noncommutative. The multiplication in it is specified by the formulas

$$[e_1, e_2] = e_3, \qquad [e_2, e_3] = e_1, \qquad [e_3, e_1] = e_2.$$

Important examples of Lie algebras are provided by tangent spaces at the identity to the groups of motions with usual operation of matrix commutation (6.4). We considered them in Section 5.2.1 and will give a detailed list of such algebras in Section 3. They are Lie subalgebras of the algebra $\mathrm{M}(n, \mathbb{R})$. This means the following.

Let $V' \subset V$ be a linear subspace of a Lie algebra $V$, which is closed with respect to the commutation operation (commutator)

$$[V', V'] \subset V'.$$

Then $V'$ itself is a Lie algebra with respect to this operation. This algebra is said to be a *Lie subalgebra* of the Lie algebra $V$.

There is also a converse relationship, which enables one to recover Lie groups from Lie algebras.

We say that we are given an $n$-dimensional *local Lie group* if there are smooth mappings

$$\varphi\colon U \times U \to \mathbb{R}^n, \qquad \psi\colon U \to \mathbb{R}^n,$$

defined in a neighborhood $U$ of the origin $0 = (0, \ldots, 0) \in \mathbb{R}^n$ such that the following relations hold:

1) $\varphi(x, 0) = \varphi(0, x) = x, \ \psi(0) = 0,$
2) $\varphi(x, \psi(x)) = \varphi(\psi(x), x) = 0,$
3) $\varphi(\varphi(x, y), z) = \varphi(x, \varphi(y, z)),$

provided that the terms involved in these relations are well defined (if $\psi(x) \in U$ in relation 2) and if $\varphi(x, y)$ and $\varphi(y, z)$ belong to $U$ in relation 3)). Relation 1) and continuity of the mappings $\varphi$ and $\psi$ imply that there exists a neighborhood of the origin $W \subset U$ such that $\varphi(x, y), \varphi(y, z) \in U$ and $\psi(x) \in U$ for $x, y, z \in W$. In this case all the terms involved in relations 1)–3) are well defined.

We see that the operations $xy = \varphi(x, y)$ and $x^{-1} = \psi(x)$ satisfy the same requirements as multiplication and inversion in a group with the identity element $1 = (0, \ldots, 0)$.

A mapping $\varphi\colon U \to V$ of local Lie groups $U$ and $V$ such that $\varphi(xy) = \varphi(x)\varphi(y)$ is called a homomorphism of these groups. Two local groups are said to be isomorphic if for some neighborhoods $U$ and $V$ of their identities there exists a one-to-one homomorphism $\varphi\colon U \to V$ of local groups: $\varphi(U) = V$, and $\varphi(x) = \varphi(y)$ in $V$ if and only if $x = y$ in $U$.

The following theorem is due to Lie.

**Theorem 6.3.** *For any finite-dimensional Lie algebra $V$ over the field of real numbers, one can construct (uniquely up to isomorphism) a local Lie group for which this algebra is the tangent algebra at the identity with commutator* (6.2).

For a Lie group $G$, any sufficiently small neighborhood $U$ of the identity element is a local Lie group. Two Lie groups $G$ and $H$ are said to be *locally isomorphic* if there are neighborhoods $U$ and $V$ of their identities such that they are isomorphic as local Lie groups.

**Corollary 6.2.** *Two Lie algebras are isomorphic if and only if the corresponding local Lie groups are isomorphic.*

For any Lie algebra there may be a family of pairwise nonisomorphic Lie groups for which it is the tangent algebra. Theorem 6.2 presents an example of this situation: groups $\mathrm{SL}(2, \mathbb{R})$ and $\widetilde{\mathrm{SL}}(2, \mathbb{R})$. Therefore, the Lie algebras provide a classification of Lie groups up to a local isomorphism.

In fact, the following result holds.

*For any Lie algebra $\mathfrak{g}$ there exists a unique simply connected (as a topological space) Lie group $G$ for which the Lie algebra $\mathfrak{g}$ is the tangent algebra. All other Lie groups locally isomorphic to $G$ have the form $G/\Gamma$, where $\Gamma$ is a discrete normal subgroup of the group $G$.*

A subgroup $F = \{x(t)\}$ of a Lie group is a *one-parameter* subgroup if it is parametrized by the points of the real line and

$$x(s)x(t) = x(s+t)$$

for all $s, t \in \mathbb{R}$. In particular, $x(0) = 1 \in G$. Note that we do not require that $x(t) \neq x(s)$ for $t \neq s$.

**Theorem 6.4.** *All the one-parameter subgroups $x(t)$ in a Lie group $G$ are solutions to the ordinary differential equation*

$$\dot{x} = x\xi$$

*with a constant vector $\xi \in T_1G$ and initial condition $x(0) = 1$.*

Here and afterwards, the product of a group element $x$ and a vector $\xi$ means the action of the differential of left multiplication by $x$ on the tangent vector $\xi$:

$$g \to xg, \qquad \xi \in T_1G \to x\xi \in T_xG.$$

**Proof.** First of all we observe that in a neighborhood of each point $x \in G$ the Lie group has the structure of a Euclidean space. Therefore, to prove the existence and uniqueness of the solution to such an equation in the Lie group we can invoke the corresponding fact for ordinary differential equations $\dot{x} = v(x)$ in Euclidean spaces.

If $x(t)$ is a one-parameter subgroup in $G$, then

$$\dot{x}(t) = \left.\frac{dx(t+\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = x(t)\left.\frac{dx(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = x(t)\dot{x}(0);$$

hence the one-parameter subgroup satisfies the ordinary differential equation

$$x^{-1}\dot{x} = \xi, \quad \xi = \dot{x}(0) \in T_1G,$$

on the Lie group.

If $V$ is the tangent space to the group $G$ at the identity, then to each vector $\xi \in V$ corresponds the curve $x(t)$, which is the solution to the ordinary differential equation

$$x^{-1}\dot{x} = \xi$$

with initial condition $x(0) = 1$ in the Lie group $G$. Fix some $s \neq 0$ and consider the curve $y(t) = x(s)x(t)$. It satisfies the differential equation

$$y^{-1}\dot{y} = \xi$$

with $y(0) = x(s)$. Note that the curve $z(t) = x(s+t)$ satisfies the same equation with the same initial condition. Therefore, these curves coincide, and we have $x(s+t) = x(s)x(t)$. Since $s$ may be chosen arbitrarily, this relation holds for all $s$ and $t$. Hence we conclude that $x(t)$ is a one-parameter subgroup. □

EXAMPLE. The one-parameter subgroups for the matrix group $\mathrm{GL}(n, \mathbb{R})$ have the form $e^{\xi t} = \exp(\xi t)$. Therefore, for arbitrary Lie groups they also have this form, and the mapping

$$\xi \to \exp(\xi),$$

which assigns to an element $\xi \in T_1 G$ of the Lie algebra the element $e^{\xi t}$ with $t = 1$ of the one-parameter subgroup, is called the *exponential mapping*.

By means of the exponential mapping one can introduce two different types of local coordinates in a neighborhood of the identity. Let $e_1, \ldots, e_n$ be a basis of the Lie algebra. Then each point $g$ in a sufficiently small neighborhood of the identity is uniquely representable as the image of the exponential mapping,

$$g = \exp(x^i e_i).$$

This follows from the fact that the Jacobian of the exponential mapping at zero is equal to the identity matrix, which can be easily verified. We take $(x^1, \ldots, x^n)$ for the coordinates of the point $g$. These are the *coordinates of the first kind.*

At the same time each point in a small neighborhood of the identity is representable as

$$g = \exp(t_1 e_1) \cdots \exp(t_n e_n).$$

The $n$-tuple $(t_1, \ldots, t_n)$ specifies coordinates of the point $g$. They are called the *coordinates of the second kind.*

By means of the exponential mapping we can give another definition of the commutator in a Lie group. Let $\xi, \eta \in T_1 G = \mathfrak{g}$. Take the Taylor expansion in $t$ of the product $e^{\xi t} e^{\eta t} e^{-\xi t} e^{-\eta t}$ about the point $t = 0$.

**Lemma 6.3.** *In local coordinates in a neighborhood of the identity* $1 = (0, \ldots, 0)$ *of the Lie group, the following equality holds:*

$$e^{\xi t} e^{\eta t} e^{-\xi t} e^{-\eta t} = [\xi, \eta] t^2 + O(t^3).$$

**Proof.** For brevity, we restrict ourselves to the case where $G$ is the matrix

group. The general case is treated similarly. For $\xi = X \in \mathrm{M}(n,\mathbb{R})$, $\eta = Y \in \mathrm{M}(n,\mathbb{R})$ we have

$$\begin{aligned} e^{Xt}e^{Yt}e^{-Xt}e^{-Yt} &= \Big(1 + Xt + \frac{1}{2}X^2t^2 + O(t^3)\Big)\Big(1 + Yt + \frac{1}{2}Y^2t^2 + O(t^3)\Big) \\ &\quad\times \Big(1 - Xt + \frac{1}{2}X^2t^2 + O(t^3)\Big)\Big(1 - Yt + \frac{1}{2}Y^2t^2 + O(t^3)\Big) \\ &= 1 + (XY - YX)t^2 + O(t^3). \end{aligned}$$

This proves the lemma. □

**Corollary 6.3.** *Let $F\colon G \to H$ be a smooth homomorphism of matrix groups. Then the differential $F_*$ of this mapping at the identity of the group is a homomorphism of Lie algebras:*

$$F_*([X,Y]) = [F_*(X), F_*(Y)].$$

**Proof.** The homomorphism $F$ of Lie groups maps one-parameter subgroups $e^{Xt}$ into one-parameter subgroups $e^{X't}$, where $X' = F_*(X)$. This was proved for matrix groups in Lemma 5.12. This implies that

$$F(e^{Xt}e^{Yt}e^{-Xt}e^{-Yt}) = e^{X't}e^{Y't}e^{-X't}e^{-Y't} = 1 + [X',Y']t^2 + O(t^3),$$

where $e^{Xt}e^{Yt}e^{-Xt}e^{-Yt} = 1 + [X,Y]t^2 + O(t^3)$. Hence we conclude that

$$F_*([X,Y]) = [F_*(X), F_*(Y)].$$

This proves the corollary. □

There is an explicit formula, written in terms of the exponential mapping, for construction of the local Lie group for a Lie algebra. We state this *Campbell–Hausdorff formula* without proof.

**Theorem 6.5.** *Let $V$ be a Lie algebra. Let the mapping $V \times V \to V$ be specified by the formula*

$$(x,y) \to z = \sum_{n=0}^{\infty} z_n, \quad z_n = \sum_{p+q=n} (z'_{p,q} + z''_{p,q}),$$

*where*

$$\begin{aligned} z'_{p,q} &= \frac{1}{p+q} \\ &\times \sum \frac{(-1)^{m+1}}{m} \frac{\operatorname{ad}(x)^{p_1}\operatorname{ad}(y)^{q_1}\cdots\operatorname{ad}(x)^{p_{m-1}}(y)\operatorname{ad}(x)^{q_{m-1}}(x)\operatorname{ad}(x)^{p_m}(y)}{p_1!\,q_1!\cdots p_{m-1}!\,q_{m-1}!\,p_m!} \end{aligned}$$

*(with summation over $m \geq 1$, $p_1 + \cdots + p_m = p, q_1 + \cdots + q_{m-1} = q - 1$, $p_i + q_i \geq 1$ for $i = 1, \ldots, m-1$, and $p_m \geq 1$) and*

$$z''_{p,q} = \frac{1}{p+q} \sum \frac{(-1)^{m+1}}{m} \frac{\operatorname{ad}(x)^{p_1}\operatorname{ad}(y)^{q_1}\cdots\operatorname{ad}(y)^{p_{m-1}}(x)\operatorname{ad}(y)^{q_{m-1}}(x)}{p_1!\,q_1!\cdots p_{m-1}!\,q_{m-1}!}$$

*(with summation over $m \geq 1$, $p_1 + \cdots + p_{m-1} = p - 1$, $q_1 + \cdots + q_{m-1} = q$, $p_i + q_i \geq 1$ for $i = 1, \ldots, m$). The mapping*

$$e^x \cdot e^y \to e^z$$

*specifies multiplication in the local Lie group, and the series for $z$ converge in a sufficiently small neighborhood of zero.*

If there exists $N$ such that

$$[\xi_1, [\xi_2, [\ldots, [\xi_N, \xi_{N+1}] \cdots]]] = 0$$

for any elements $\xi_1, \ldots, \xi_{N+1}$ of the Lie algebra $V$, then $V$ is said to be a *nilpotent Lie algebra* of class (degree) $N$. In this case the series for $z$ is a polynomial of degree $N$, and we obtain a polynomial multiplication rule $z = P(x, y)$ in the Lie group. For example, the Heisenberg group is a nilpotent Lie group of degree two with multiplication rule

$$(x, y, z) \cdot (x', y', z') = (x + x', y + y', z + z' + xy').$$

Note that, unlike Lie groups, all Lie algebras are matrix algebras. This is established by the following Ado theorem.

**Theorem 6.6.** *If $V$ is a finite-dimensional algebra over the field of real or complex numbers, then there exists its linear imbedding*

$$\varphi\colon V \to \mathrm{M}(n, \mathbb{R}) \quad \text{or} \quad \mathrm{M}(n, \mathbb{C})$$

*into the algebra of square matrices over the same field, which is a homomorphism of Lie algebras:*

$$\varphi([A, B]) = \varphi(A)\varphi(B) - \varphi(B)\varphi(A).$$

**6.1.3. Main matrix groups and Lie algebras.** Here we list the most important examples of matrix Lie groups (subgroups of the Lie group $\mathrm{GL}(n, \mathbb{R})$) and their Lie algebras.

We denote the groups by capital letters and their tangent spaces at the identity (*tangent algebras*) by the corresponding small German letters.

1. The special linear group $\mathrm{SL}(n, \mathbb{R})$ (or $\mathrm{SL}(n, \mathbb{C})$) consists of all real (complex) matrices of order $n$ with determinant equal to 1. The tangent space $\mathfrak{sl}(n, \mathbb{R})$ (or $\mathfrak{sl}(n, \mathbb{C})$) at the identity is the space of zero-trace matrices (Theorem 5.15).

2. The group of rotations $\mathrm{SO}(n, \mathbb{R})$ (or $\mathrm{SO}(n, \mathbb{C})$) consists of all real (complex) orthogonal matrices with determinant 1:

$$A^\top A = 1, \quad \det A = 1, \quad A \in \mathrm{SO}(n, \mathbb{R}),\ \mathrm{SO}(n, \mathbb{C}).$$

By Theorem 5.16 the tangent algebra $\mathfrak{so}(n, \mathbb{R})$ ($\mathfrak{so}(n, \mathbb{C})$) is the algebra of skew-symmetric real (complex) matrices of order $n$:

$$X^\top = -X, \quad X \in \mathfrak{so}(n, \mathbb{R}),\ \mathfrak{so}(n, \mathbb{C}). \tag{6.10}$$

3. Pseudoorthogonal groups $\mathrm{SO}(p, q)$. If $G = (g_{ij})$ is a pseudo-Euclidean metric in the space $\mathbb{R}^n_{p,q}$, $p + q = n$, then the group $\mathrm{SO}(p, q)$ consists of all real matrices $A$ with determinant 1 preserving the form $G = (g_{ij})$:

$$A^\top GA = G, \quad \det A = 1, \quad A \in \mathrm{SO}(p, q).$$

The tangent algebra $\mathfrak{so}(p, q)$ is the algebra of matrices $X = (x^i_j)$ such that

$$GX + X^\top G = 0 \tag{6.11}$$

(Theorem 5.17). This equality implies that $GX$ is a skew-symmetric matrix. Therefore, the mapping

$$X \to GX$$

specifies an isomorphism of linear spaces $\mathfrak{so}(p, q)$ and $\mathfrak{so}(p+q)$. This isomorphism does not preserve the commutator, and hence is not an isomorphism of Lie algebras.

4. The symplectic group $\mathrm{Sp}(n, \mathbb{R})$ consists of matrices of order $2n$ preserving a nondegenerate skew-symmetric scalar product on $\mathbb{R}^{2n}$. By Theorem 2.1 such a scalar product is specified in an appropriate basis by the matrix

$$G = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{pmatrix},$$

hence we have

$$A^\top GA = G, \quad A \in \mathrm{Sp}(n, \mathbb{R}).$$

The algebra $\mathfrak{sp}(n, \mathbb{R})$ consists of all matrices $X$ of order $2n$ satisfying the condition

$$X^\top G + GX = 0.$$

Since $G^\top = -G$, this means that the matrix $GX$ is symmetric.

The group $\mathrm{Sp}(n, \mathbb{C})$ is defined as the group of linear transformations of $\mathbb{C}^{2n}$ preserving the symplectic form specified by the matrix $G$.

5. The unitary group $\mathrm{U}(n)$ consists of matrices of order $n$ with the unitary property

$$A^\top \bar{A} = 1, \quad A \in \mathrm{U}(n).$$

Its tangent algebra $\mathfrak{u}(n)$ consists of skew-Hermitian matrices

$$X^\top = -\bar{X}, \quad X \in \mathfrak{u}(n) \tag{6.12}$$

(Theorem 5.20).

6. The special unitary group $\mathrm{SU}(n)$ is the group of unitary matrices with determinant 1. Its tangent algebra $\mathfrak{su}(n)$ is the intersection of the algebras $\mathfrak{sl}(n,\mathbb{C})$ and $\mathfrak{u}(n)$, and so $\mathfrak{su}(n)$ consists of zero-trace skew-Hermitian matrices

$$X^\top = -\bar{X}, \quad \operatorname{Tr} X = 0, \quad X \in \mathfrak{su}(n).$$

7. The pseudounitary group $\mathrm{U}(p,q)$ consists of linear transformations of $\mathbb{C}^{p+q}$ preserving the pseudo-Hermitian scalar product

$$\langle \xi, \eta \rangle = \sum_{i=1}^{p} x^i \bar{y}^i - \sum_{i=p+1}^{n} x^i \bar{y}^i = g_{ij} x^i \bar{y}^i, \tag{6.13}$$

$$\xi = (x^1, \ldots, x^n), \qquad \eta = (y^1, \ldots, y^n), \quad n = p + q.$$

If the matrix $G = (g_{ij})$ specifies the form (6.13), then the matrix $A \in \mathrm{U}(p,q)$ satisfies the equation

$$A^\top G \bar{A} = G, \quad A \in \mathrm{U}(p,q).$$

The algebra $\mathfrak{u}(p,q)$ consists of matrices $X$ such that

$$G\bar{X} + X^\top G = 0. \tag{6.14}$$

Similarly to the real case the mapping $X \to GX$ determines an isomorphism of the linear space $\mathfrak{u}(p,q)$ onto the space of skew-Hermitian matrices.

8. The group $\mathrm{SU}(p,q)$ is the subgroup of the group $\mathrm{U}(p,q)$ consisting of matrices with determinant 1, and the algebra $\mathfrak{su}(p,q)$ is the subalgebra of $\mathfrak{u}(p,q)$ formed by the zero-trace matrices, $\operatorname{Tr} X = 0$.

9. The group of quaternion-linear transformations $\mathrm{Sp}(n)$ preserving the Hermitian scalar product on $\mathbb{H}^n$ is imbedded into $\mathrm{GL}(2n,\mathbb{C})$ as the intersection

$$\mathrm{Sp}(n) = \mathrm{Sp}(n,\mathbb{C}) \cap \mathrm{U}(2n)$$

(see Section 5.3.3).

10. The affine group $\mathrm{A}(n)$ of all affine transformations

$$x \to Ax + b, \quad A \in \mathrm{GL}(n), \quad b \in \mathbb{R}^n,$$

of the $n$-dimensional Cartesian space and its subgroup $\mathrm{E}(n)$ formed by all motions of this space and specified in $\mathrm{A}(n)$ by the condition $A \in \mathrm{O}(n) \subset \mathrm{GL}(n)$.

We will show that the tangent algebras of the matrix groups listed above are Lie subalgebras of the Lie algebras $\mathrm{M}(n,\mathbb{R})$ or $\mathrm{M}(n,\mathbb{C})$.

**Theorem 6.7.** *The linear spaces $\mathfrak{sl}(n,\mathbb{R})$, $\mathfrak{sl}(n,\mathbb{C})$, $\mathfrak{so}(n,\mathbb{R})$, $\mathfrak{so}(n,\mathbb{C})$, $\mathfrak{so}(p,q)$, $\mathfrak{sp}(n,\mathbb{R})$, $\mathfrak{sp}(n,\mathbb{C})$, $\mathfrak{u}(n)$, $\mathfrak{su}(n)$, $\mathfrak{u}(p,q)$, $\mathfrak{su}(p,q)$, $\mathfrak{sp}(n)$, $\mathfrak{a}(n)$, and $\mathfrak{e}(n)$ are Lie algebras relative to the matrix commutation* (6.4).

**Proof.** Since $\operatorname{Tr} XY = \operatorname{Tr} YX$, we obtain $\operatorname{Tr}[X,Y] = 0$; hence the spaces $\mathfrak{sl}(n,\mathbb{R})$ and $\mathfrak{sl}(n,\mathbb{C})$ are closed with respect to commutation and are Lie algebras.

Let us show that whenever matrices $X$ and $Y$ satisfy condition (6.14) for a real symmetric matrix $G$, the commutator of these matrices also satisfies this condition. For real matrices condition (6.14) takes the form (6.11). For $G = 1$ we obtain condition (6.10) for real matrices and condition (6.12) for complex matrices.

Suppose matrices $X$, $Y$ satisfy condition (6.11):

$$X^\top G = -G\bar{X}, \qquad Y^\top G = -G\bar{Y}.$$

Then

$$\begin{aligned}[X,Y]^\top G &= Y^\top X^\top G - X^\top Y^\top G = -Y^\top G\bar{X} + X^\top G\bar{Y} \\ &= G\bar{Y}\bar{X} - G\bar{X}\bar{Y} = -G\overline{[X,Y]},\end{aligned}$$

which proves the theorem. □

Now we arrive at the following important notion. Let $G$ be one of the matrix transformation groups listed in 1–10. The tangent space at its identity endowed with the operation of matrix commutation is called the *Lie algebra of the group* $G$ and is denoted by $\mathfrak{g}$.

EXAMPLES. 1. The Lie algebra $\mathfrak{so}(3,\mathbb{R})$ of the group of rotations of the three-dimensional space consists of skew-symmetric $3 \times 3$ matrices. We introduce the following basis $e_1, e_2, e_3$ in this space:
(6.15)

$$e_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad e_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The commutator is specified by the formulas

$$[e_1, e_2] = e_3, \qquad [e_2, e_3] = e_1, \qquad [e_3, e_1] = e_2.$$

**Corollary 6.4.** *The Lie algebra $\mathfrak{so}(3,\mathbb{R})$ is isomorphic to the algebra of vectors of the three-dimensional Euclidean space with vector product operation.*

2. The Lie algebra $\mathfrak{su}(2)$. Take the following basis $s_1, s_2, s_3$ in this algebra:

$$s_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad s_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad s_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \tag{6.16}$$

The commutation relations are

$$[s_1, s_2] = 2s_3, \qquad [s_2, s_3] = 2s_1, \qquad [s_3, s_1] = 2s_2.$$

**Theorem 6.8.** *There is a Lie algebra isomorphism*

$$\mathfrak{su}(2) \xrightarrow{\simeq} \mathfrak{so}(3, \mathbb{R}).$$

**Proof.** This isomorphism has a particularly simple form in the bases (6.15) and (6.16) for the algebras $\mathfrak{so}(3, \mathbb{R})$ and $\mathfrak{su}(2)$:

$$s_i \to 2e_i, \quad i = 1, 2, 3.$$

It is specified as the isomorphism of tangent spaces at the identity induced by the smooth isomorphism of the matrix groups

$$\mathrm{SU}(2)/\pm 1 \to \mathrm{SO}(3)$$

given by (5.29). □

In a similar way, Theorem 6.8 and the existence of the smooth isomorphism of the groups $(\mathrm{SU}(2) \times \mathrm{SU}(2))/\pm 1$ and $\mathrm{SO}(4)$ given by (5.30) imply the following result.

**Theorem 6.9.** *There is a Lie algebra isomorphism*

$$\mathfrak{so}(3, \mathbb{R}) \times \mathfrak{so}(3, \mathbb{R}) \xrightarrow{\simeq} \mathfrak{so}(4, \mathbb{R}).$$

3. The Lie algebra $\mathfrak{sl}(2, \mathbb{R})$. In the basis $Y_0, Y_1, Y_2$ defined by

$$(6.17) \qquad Y_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad Y_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad Y_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

the commutators become

$$(6.18) \qquad [Y_0, Y_1] = -2Y_2, \qquad [Y_0, Y_2] = 2Y_1, \qquad [Y_1, Y_2] = 2Y_0.$$

**Theorem 6.10.** *There is a Lie algebra isomorphism*

$$\mathfrak{sl}(2, \mathbb{R}) \xrightarrow{\simeq} \mathfrak{so}(1, 2).$$

**Proof.** Define the quadratic form

$$(6.19) \qquad \langle Y, Y \rangle = \det Y$$

on the linear space $\mathfrak{sl}(2, \mathbb{R})$. This is a pseudo-Euclidean form of signature $(1, 2)$, since

$$\det Y = y_0^2 - y_1^2 - y_2^2$$

for $Y = y_0 Y_0 + y_1 Y_1 + y_2 Y_2$.

With each transformation $A \in \mathrm{SL}(2, \mathbb{R})$ we associate the linear transformation

$$Y \to AYA^{-1}$$

of the space $\mathfrak{sl}(2, \mathbb{R})$. It preserves the linear form (6.19), since $\det Y = \det AYA^{-1}$; hence it specifies a smooth homomorphism of matrix groups

$$F \colon \mathrm{SL}(2, \mathbb{R}) \to \mathrm{SO}(1, 2)$$

similar to homomorphisms (5.29) and (5.30). The mapping $F$ induces a linear mapping of tangent spaces at the identity:

$$F_*\colon \mathfrak{sl}(2,\mathbb{R}) \to \mathfrak{so}(1,2).$$

By definition, this mapping assigns to each matrix in $\mathfrak{sl}(2,\mathbb{R})$ the linear operator on $\mathfrak{sl}(2,\mathbb{R})$ that acts by the formula

$$\begin{aligned} F_*(X)(Y) &= \frac{d}{dt}\Big|_{t=0}(e^{Xt}Ye^{-Xt}) \\ &= \frac{d}{dt}\Big|_{t=0}[(1+Xt+o(t))Y(1-Xt+o(t))] = [X,Y] = \operatorname{ad}X(Y). \end{aligned}$$

Thus we have shown that

$$F_*(X) = \operatorname{ad}X.$$

Since

$$\langle F(e^{Xt})(Y), F(e^{Xt})(Y)\rangle = \langle Y,Y\rangle,$$

by differentiating both sides of this equality with respect to $t$ at $t=0$ we obtain

$$\langle \operatorname{ad}(X)(Y), Y\rangle = -\langle Y, \operatorname{ad}(X)(Y)\rangle,$$

i.e., the operators $\operatorname{ad}X$ are skew-symmetric relative to the metric (6.19).

The commutation relations (6.18) imply that there is no nonzero matrix $X$ such that $\operatorname{ad}X = 0$. Therefore, the mapping $F_*$ has zero kernel and, since the groups $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{SO}(1,2)$ have the same dimension, is a Lie algebra homomorphism. This completes the proof. □

4. The Lie algebra $\mathfrak{a}(n)$ of the affine group $\mathrm{A}(n)$ consists of $n\times n$ matrices of the form

$$\begin{pmatrix} X & Y \\ 0 & 0 \end{pmatrix}, \quad X \in \mathfrak{gl}(n), \quad Y \in \mathbb{R}^n.$$

The commutator is

$$\begin{aligned} [(X_1,Y_1),(X_2,Y_2)] &= (X_1X_2 - X_2X_1, X_1Y_2 - X_2Y_1) \\ &= ([X_1,X_2],(X_1Y_2 - X_2Y_1)). \end{aligned}$$

Here $[X_1,X_2]$ is the usual matrix commutator of order $n$, and $XY$ denotes the vector obtained by the action of the matrix $X$ on the vector $Y$.

The Lie algebra $\mathfrak{e}(n)$ of the group of motions $\mathrm{E}(n)$ is the subalgebra in $\mathfrak{a}(n)$ specified by the condition $X \in \mathfrak{so}(n)$.

For example, the Lie algebra of the group $\mathrm{E}(3)$ of motions of the three-dimensional Euclidean space is generated by the matrices $e_1, e_2, e_3$ in (6.15) and the vectors $f_1 = (1,0,0)^\top$, $f_2 = (0,1,0)^\top$, $f_3 = (0,0,1)^\top$. The commutation relations are

$$[e_i,e_j] = \varepsilon_{ijk}e_k, \quad [e_i,f_j] = \varepsilon_{ijk}f_k, \quad [f_i,f_j] = 0, \qquad i,j = 1,2,3,$$

where

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } \left(\begin{smallmatrix} 1 & 2 & 3 \\ i & j & k \end{smallmatrix}\right) \text{ is an even permutation,} \\ -1 & \text{if } \left(\begin{smallmatrix} 1 & 2 & 3 \\ i & j & k \end{smallmatrix}\right) \text{ is an odd permutation,} \\ 0 & \text{if } i, j, k \text{ are not all different,} \end{cases} \tag{6.20}$$

and summation over $k$ is assumed.

5. The Lie algebra $\mathfrak{g}$ of the Heisenberg group of matrices of the form

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}, \quad x, y, z \in \mathbb{R},$$

with usual multiplication. This Lie algebra is generated by the matrices

$$e_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with commutation relations

$$[e_1, e_2] = e_3, \qquad [e_1, e_3] = [e_2, e_3] = 0.$$

These relations imply that $[\xi, [\eta, \zeta]] = 0$ for any triple $\xi, \eta, \zeta \in \mathfrak{g}$. This means that the Lie algebra of the Heisenberg group is a nilpotent Lie algebra of degree 2.

**6.1.4. Invariant metrics on Lie groups.** Let $G$ be a Lie group or, for simplicity, one of the matrix transformation groups listed above. For any element $g \in G$, left and right multiplications by $g$ specify smooth invertible mappings $G \to G$ (with multiplication by $g^{-1}$ being the inverse mapping):

$$h \to gh, \qquad h \to hg.$$

These mappings are called the *left* and *right translations* by an element $g \in G$. Obviously, the left translations commute with the right translations.

Being smooth invertible mappings, translations induce isomorphisms of tangent spaces,

$$T_h G \to T_{gh} G, \quad T_h G \to T_{hg} G.$$

For matrix groups these isomorphisms have a very simple form: if $X \in \mathrm{M}(n, \mathbb{R})$ and $A \in G$, then

$$X \to AX \text{ (left translation)}, \qquad X \to XA \text{ (right translation)}.$$

For simplicity, we will use this notation for translations in all groups.

The composition of the left translation by $g$ and the right translation by $g^{-1}$ determines an automorphism of the tangent Lie algebra $\mathfrak{g} = T_1 G$:

$$\operatorname{Ad} g \colon \mathfrak{g} \to \mathfrak{g}, \quad \operatorname{Ad} g(\xi) = g \xi g^{-1}.$$

Recall that a finite-dimensional *representation* of an arbitrary group $G$ is a homomorphism $\rho$ of this group into the linear group $\mathrm{GL}(n,\mathbb{R})$:

$$\rho\colon G \to \mathrm{GL}(n,\mathbb{R}).$$

A representation is said to be *exact* if this homomorphism is injective: $\rho(g) \neq \rho(h)$ for any pair of different elements $g, h \in G$. A representation is said to be *irreducible* if it has no nontrivial invariant subspaces. This means that if $V \subset \mathbb{R}^n$ is a linear subspace and $\rho(G)V \subset V$, then either $V = 0$ or $V = \mathbb{R}^n$.

We see that the mapping

$$\mathrm{Ad}\colon G \to \mathrm{GL}(n,\mathbb{R}), \quad n = \dim \mathfrak{g},$$

is a representation of the Lie group $G$. It is called the *adjoined representation* of the group $G$.

**Theorem 6.11.** *The differential of the mapping* $\mathrm{Ad}\colon G \to \mathrm{GL}(n,\mathbb{R})$ *at the identity of the Lie group is the mapping*

$$\mathrm{ad}\colon \mathfrak{g} \to \mathfrak{gl}(n,\mathbb{R}), \quad \mathrm{ad}(X)(A) = [X, A],$$

*where* $X, A \in \mathfrak{g}$ *and the operator* $\mathrm{ad}(X)$ *acts linearly on the space* $\mathbb{R}^n$*, which is isomorphic as a linear space to the Lie algebra* $\mathfrak{g}$.

**Proof.** We will use the Taylor expansion of multiplication in the Lie group and for simplicity present the proof for matrix groups. It suffices to consider $\mathrm{GL}(N,\mathbb{R})$. We have $\mathrm{Ad}\,1 = 1$ and

$$\begin{aligned} \mathrm{Ad}\, e^{tX}(A) = e^{tX} A e^{-tX} &= (1 + tX + O(t^2))A(1 - tX + O(t^2)) \\ &= A + (XA - AX)t + O(t^2) = A + [X, A]t + O(t^2). \end{aligned}$$

This proves the theorem. □

Let $\langle \cdot, \cdot \rangle_0$ be a scalar product in the Lie algebra $\mathfrak{g} = T_1 G$ of the Lie group $G$. It generates the metrics on the entire group $G$ defined by the formulas

$$\langle \xi, \eta \rangle_L = \langle g^{-1}\xi, g^{-1}\eta \rangle_0, \qquad \langle \xi, \eta \rangle_R = \langle \xi g^{-1}, \eta g^{-1} \rangle_0 \tag{6.21}$$

for $\xi, \eta \in T_g G$.

The metric $\langle \cdot, \cdot \rangle_L$ is invariant with respect to left translations (is *left-invariant)*:

$$\langle \xi, \eta \rangle_L = \langle g\xi, g\eta \rangle_L.$$

Indeed, we have

$$\langle h\xi, h\eta \rangle_L = \langle (hg)^{-1} h\xi, (hg)^{-1} h\eta \rangle_0 = \langle g^{-1}\xi, g^{-1}\eta \rangle_0 = \langle \xi, \eta \rangle_L$$

for all $\xi, \eta \in T_g G$. It can be shown in a similar way that $\langle \xi, \eta \rangle_R$ is a *right-invariant* metric: $\langle \xi, \eta \rangle_R = \langle \xi g, \eta g \rangle_R$.

**Theorem 6.12.** *Any invariant metric on a Lie group $G$ is determined by the scalar product on the Lie algebra $\mathfrak{g}$ according to* (6.21).

**Proof.** For definiteness, let $\langle\cdot,\cdot\rangle$ be a left-invariant metric. Consider its restriction to the tangent space at the identity:

$$\langle\xi,\eta\rangle_0 = \langle\xi,\eta\rangle, \quad \xi,\eta \in T_1G = \mathfrak{g}.$$

It follows from the left invariance that

$$\langle g\xi, g\eta\rangle = \langle\xi,\eta\rangle_0,$$

i.e., this metric satisfies (6.21). □

A metric in a group $G$ is said to be bilaterally invariant if it is invariant relative to both left and right translations.

**Theorem 6.13.** *Let a bilaterally invariant metric in a Lie group be specified at the identity of the group by a scalar product $\langle Y,Z\rangle$. Then the operators* $\operatorname{ad} X$ *are skew-symmetric with respect to this scalar product,*

$$\langle \operatorname{ad} X(Y), Z\rangle = -\langle Y, \operatorname{ad} X(Z)\rangle. \tag{6.22}$$

**Proof.** Obviously, a bilaterally invariant metric satisfies the equation

$$\langle \operatorname{Ad} e^{tX}(Y), \operatorname{Ad} e^{tX}(Z)\rangle = \langle Y, Z\rangle$$

for all $t$. We write down the series expansion in $t$ of the left-hand side of this equality:

$$\langle \operatorname{Ad} e^{tX}(Y), \operatorname{Ad} e^{tX}(Z)\rangle = \langle Y,Z\rangle + (\langle \operatorname{ad} X(Y), Z\rangle + \langle Y, \operatorname{ad} X(Z)\rangle)t + O(t^2).$$

Comparing these two equalities we see that

$$\langle \operatorname{ad} X(Y), Z\rangle + \langle Y, \operatorname{ad} X(Z)\rangle = 0,$$

which proves the theorem. □

Now we arrive at the following important definition. A symmetric scalar product (Euclidean or pseudo-Euclidean) in a Lie algebra $\mathfrak{g}$ is called a *Killing metric* if the operators $\operatorname{ad} X$ are skew-symmetric with respect to this scalar product for all $X \in \mathfrak{g}$, i.e., equality (6.22) holds for any vectors $X, Y, Z \in \mathfrak{g}$.

If the scalar product $\langle X, Y\rangle_0$ in the Lie algebra $\mathfrak{g}$ of a group $G$ is a Killing metric, then the left-invariant metric in the group $G$ constructed by formula (6.21) is called the *Killing metric* in the group $G$.

EXAMPLES. 1. The usual scalar product

$$\langle Y, Z\rangle = \sum_{i=1}^{3} Y^i Z^i$$

in $\mathbb{R}^3$ is a Killing metric if we regard the space $\mathbb{R}^3$ as a Lie algebra with vector product operation (this Lie algebra is isomorphic to $\mathfrak{su}(2)$).

2. The scalar product $\langle Y, Y\rangle = \det Y$ in the algebra $\mathfrak{sl}(2,\mathbb{R})$ is a pseudo-Euclidean Killing metric.

These two metrics exemplify the general construction, which is given by the following theorem.

**Theorem 6.14.** *Let $\mathfrak{g}$ be a Lie algebra. Then the scalar product*

$$\langle X, Y\rangle = -\operatorname{Tr}(\operatorname{ad} X \cdot \operatorname{ad} Y), \quad X, Y \in \mathfrak{g}, \tag{6.23}$$

*is a Killing metric in the algebra $\mathfrak{g}$.*

**Proof.** Note that the trace does not depend on the choice of a basis in the Lie algebra: this well-known fact of linear algebra is valid for operators on an arbitrary vector space $V$. Moreover, it is well known that

$$\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$$

for arbitrary operators $A, B\colon V \to V$. Hence the scalar product (6.23) is symmetric and well defined regardless of the choice of a basis in $\mathfrak{g}$.

The Jacobi identity can be rewritten as

$$\operatorname{ad}[X,Y] = \operatorname{ad} X \operatorname{ad} Y - \operatorname{ad} Y \operatorname{ad} X.$$

Using this equality we obtain

$$\begin{aligned}
-\langle \operatorname{ad} X(Y), Z\rangle &= \operatorname{Tr}(\operatorname{ad}[X,Y]\operatorname{ad} Z) = \operatorname{Tr}(\operatorname{ad} X \operatorname{ad} Y \operatorname{ad} Z - \operatorname{ad} Y \operatorname{ad} X \operatorname{ad} Z)\\
&= \operatorname{Tr}(\operatorname{ad} Y \operatorname{ad} Z \operatorname{ad} X - \operatorname{ad} Y \operatorname{ad} X \operatorname{ad} Z) = \operatorname{Tr}(\operatorname{ad} Y \operatorname{ad}[Z,X])\\
&= -\operatorname{Tr}(\operatorname{ad} Y \operatorname{ad}[X,Z]) = \langle Y, \operatorname{ad} X(Z)\rangle.
\end{aligned}$$

Hence the theorem. □

Theorem 6.13 also provides elegant examples of Killing metrics.

**Lemma 6.4.** 1. *The metric induced by the imbedding* $\mathrm{SO}(n,\mathbb{R}) \subset \mathrm{M}(n,\mathbb{R}) = \mathbb{R}^{n^2}$ *is bilaterally invariant.*

2. *The metric induced by the imbedding* $\mathrm{U}(n) \subset \mathrm{M}(n,\mathbb{C}) = \mathbb{R}^{2n^2}$ *is bilaterally invariant.*

**Proof.** The Euclidean metric in the space $\mathbb{R}^{n^2}$ of $n \times n$ matrices has the form

$$\langle X, Y\rangle = \operatorname{Tr}(XY^\top).$$

Let $X, Y \in \mathfrak{so}(n,\mathbb{R})$ and $A \in \mathrm{SO}(n,\mathbb{R})$. Since the metric in $\mathrm{SO}(n,\mathbb{R})$ is induced by the imbedding, we have

$$\langle AX, AY\rangle = \operatorname{Tr}(AX(AY)^\top) = \operatorname{Tr}(AXY^\top A^\top).$$

Interchanging the factors does not affect the trace and $A^\top A = 1$; hence

$$\operatorname{Tr}(AXY^\top A^\top) = \operatorname{Tr}(A^\top AXY^\top) = \operatorname{Tr}(XY^\top) = \langle X, Y\rangle.$$

Therefore, the induced metric is left-invariant. Moreover,

$$\langle XA, YA\rangle = \operatorname{Tr}(XA(YA)^\top) = \operatorname{Tr}(XAA^\top Y) = \operatorname{Tr}(XY) = \langle X, Y\rangle.$$

This shows that the induced metric in $\mathrm{SO}(n,\mathbb{R})$ is right-invariant as well.

The proof for the metric in $\mathrm{U}(n)$ is similar. Note only that the Euclidean metric in $\mathrm{M}(n,\mathbb{C})$ is

$$\langle X, Y\rangle = \operatorname{Re}\operatorname{Tr}(X\bar{Y}^\top);$$

hence the induced metric in the Lie algebra has the form

$$\langle X, Y\rangle = \operatorname{Re}\operatorname{Tr}(X\bar{Y}^\top) = -\operatorname{Re}\operatorname{Tr} XY.$$

The proof is completed. □

**Corollary 6.5.** *The metrics of the groups* $\mathrm{SO}(n,\mathbb{R})$ *and* $\mathrm{U}(n)$ *induced by their imbeddings into* $\mathrm{M}(n,\mathbb{R})$ *and* $\mathrm{M}(n,\mathbb{C})$ *are Killing's metrics.*

**6.1.5. Homogeneous spaces.** Here we point out an important class of spaces related to Lie groups.

**Theorem 6.15.** *Let $G$ be a Lie group, and $H$ a closed Lie subgroup of it. Then on the space $G/H$ of left cosets by the subgroup $H$, one can introduce a structure of smooth manifold of dimension*

$$\dim G/H = \dim G - \dim H$$

*such that the projection*

$$\pi\colon G \to G/H$$

*is a smooth mapping.*

**Proof.** The closed subgroup $H$ acts on $G$ by translations

$$g \to gh^{-1}, \quad g \in G,\ h \in H.$$

If this action is discrete, then $\dim H = 0$ and $H$ is said to be a *discrete subgroup*. For this case we have constructed the required smooth structure on the quotient space $G/H$ in Theorem 5.8.

Suppose that $\dim H \geq 1$. Decompose the Lie algebra $\mathfrak{g}$ into the direct sum

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{h},$$

where $\mathfrak{h}$ is the Lie algebra of the group $H$ and $\mathfrak{k}$ is the linear complement of $\mathfrak{h}$ in $\mathfrak{g}$. Take a basis $e_1, \dots, e_m, e_{m+1}, \dots, e_n$ in $\mathfrak{g}$ such that $e_1, \dots, e_m$ form a basis in $\mathfrak{k}$ and $e_{m+1}, \dots, e_n$ a basis in $\mathfrak{h}$. In a neighborhood $U$ of

the identity of the group $G$ we construct the coordinates of the second kind corresponding to this basis,

$$g = e^{t_1 e_1} \dots e^{t_n e_n} \leftrightarrow (t_1, \dots, t_n).$$

Next, we find a neighborhood $V \subset U$ of $1 \in G$ such that $V \cdot V \subset U$.

If $g_1 = (r_1, \dots, r_m, 0, \dots, 0)$ and $g_2 = (s_1, \dots, s_m, 0, \dots, 0)$ belong to $V$, then the left cosets $g_1 H$ and $g_2 H$ coincide if and only if $r_i = s_i$, $i = 1, \dots, m$. Indeed, if $g_1 = g_2 h$, where $h \in H$, then $h = g_2^{-1} g_1 \in U$ and we have

$$g_1 = e^{r_1 e_1} \dots e^{r_m e_m} = e^{s_1 e_1} \dots e^{s_m e_m} e^{t_1 e_{m+1}} \dots e^{t_{n-m} e_n},$$

where $h = e^{t_1 e_{m+1}} \dots e^{t_{n-m} e_n}$. But by the definition of coordinates of the second kind, this is possible only for $t_1 = \dots = t_{n-m} = 0$, $r_i = s_i$, $i = 1, \dots, m$.

Hence for a given $V$ we can construct a domain $V'$ in $G/H$ by assigning a point $y \in G/H$ to $V'$ if it is the orbit of a point $x \in M$ with coordinates $(x_1, \dots, x_m, 0, \dots, 0)$, and we take the tuple $(x_1, \dots, x_m)$ for the coordinates of the point $y \in G/H$. In this way we construct smooth neighborhoods of a coset $H \in G/H$ (the orbit of the unit of action of the group $H$).

Let $gH \in G/H$ and let $V'$ be a neighborhood of $H$ in $G/H$. Then we regard the totality of cosets of the form $gV'$ as a neighborhood of $gH$ and we take for smooth coordinates in this neighborhood the coordinates which come from $V'$:

$$(gg')H = (x_1, \dots, x_m) \leftrightarrow g'H = (x_1, \dots, x_m), \quad g' \in V'.$$

As in the proof of Theorem 5.8, we now prove that the transition functions which relate different coordinate systems on the intersections of coordinate domains are smooth. Obviously, the projection $G \to G/H$ is smooth relative to this smooth structure. The proof is completed. □

There is also another definition of such spaces $G/H$.

We say that $M^n$ is a *homogeneous space* if on this space there is a smooth (left or right) action of a Lie group $G$ and this action is *transitive*, i.e., for each pair of points $x$ and $y$ in $M^n$ there exists an element $g$ of the group $G$ which carries $x$ into $y$.

The group $G$ acts on itself by left translations, and such a space is called the left principal homogeneous space of the group $G$. In a similar way the right principal homogeneous space is defined.

For definiteness, we consider the left actions of the group. All the proofs for the right actions are similar.

To each point $x \in M^n$ corresponds its *isotropy group* $H_x$ (or the *stationary group*) that consists of all elements leaving the point $x$ fixed.

**Lemma 6.5.** *All the stationary groups of points of a homogeneous space are pairwise isomorphic.*

**Proof.** Let $x, y \in M^n$ and let $g(x) = y$. Then $H_y = gH_xg^{-1}$, which proves the lemma. □

**Corollary 6.6.** *There is a one-to-one correspondence between the points of a homogeneous space $M^n$ and left cosets $G/H$, where $H$ is the isotropy group.*

We see that the homogeneous space $M^n$ and the space of left cosets $G/H$ constructed in Theorem 6.15 are diffeomorphic.

A scalar product (e.g., a Riemannian metric or a symplectic product) in a homogeneous space $M^n = G/H$ is said to be *invariant* if it is preserved under the action of the group $G$:

$$\langle \xi, \eta \rangle = \langle g_*\xi, g_*\eta \rangle$$

for all $\xi, \eta \in T_xM^n, g \in G$. It is called left-invariant or right-invariant depending on the type of the action.

EXAMPLES OF HOMOGENEOUS SPACES. 1. *The sphere* $S^n$ is specified in the Euclidean space $\mathbb{R}^{n+1}$ by the equation

$$(x^1)^2 + \cdots + (x^{n+1})^2 = 1,$$

which is fulfilled for vectors of unit length. Linear actions of the groups $\mathrm{O}(n+1)$ and $\mathrm{SO}(n+1)$ on $\mathbb{R}^{n+1}$ generate transitive actions of these groups on the sphere $S^n$: $x \to Ax$, where $x \in S^n$ and $A \in \mathrm{O}(n+1)$ or $\mathrm{SO}(n+1)$. The isotropy group of the point $(1, 0, \ldots, 0)$ consists of all block matrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}, \quad A \in \mathrm{O}(n+1) \quad \text{or} \quad \mathrm{SO}(n+1).$$

Hence we conclude that

$$S^n = \mathrm{O}(n+1)/\mathrm{O}(n) = \mathrm{SO}(n+1)/\mathrm{SO}(n).$$

If the sphere is of odd dimension $2n+1$, then it may be specified as the set of vectors of unit length in the space $\mathbb{C}^{n+1}$:

$$|z^1|^2 + \cdots + |z^{n+1}|^2 = 1.$$

In this case the groups $\mathrm{U}(n+1)$ and $\mathrm{SU}(n+1)$ act linearly on the sphere. Similarly to the real case, the isotropy groups of the point $(1, 0, \ldots, 0)$ consist of block matrices and are isomorphic to $\mathrm{U}(n)$ and $\mathrm{SU}(n)$. Therefore

$$S^{2n+1} = \mathrm{U}(n+1)/\mathrm{U}(n) = \mathrm{SU}(n+1)/\mathrm{SU}(n).$$

2. The *real projective space* $\mathbb{R}P^n$ consists of all straight lines in $\mathbb{R}^{n+1}$ passing through the origin. Similarly to the case of the sphere, the group

$\mathrm{O}(n+1)$ acts on this space linearly and transitively: an element $A \in \mathrm{O}(n+1)$ carries the line with direction vector $\xi$ into the line with direction vector $A\xi$. The isotropy group of the line with direction $(1, 0, \dots, 0)$ consists of matrices of the form

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & A \end{pmatrix}, \quad A \in \mathrm{O}(n),$$

and we conclude that

$$\mathbb{R}P^n = \mathrm{O}(n+1)/(\mathrm{O}(1) \times \mathrm{O}(n)).$$

3. The *complex projective space* $\mathbb{C}P^n$ parametrizes the subspaces of complex dimension one in $\mathbb{C}^{n+1}$. Such subspaces are specified by the direction vector $\xi$ and consist of all vectors of the form $\lambda\xi$, where $\lambda \in \mathbb{C}$. We may assume that the direction vector $\xi$ is normalized so that $|\xi| = 1$. On the space of unit direction vectors, the group $\mathrm{U}(n+1)$ acts linearly and transitively. The isotropy group of the complex line with direction vector $\xi = (1, 0, \dots, 0)$ consists of all matrices of the form

$$\begin{pmatrix} \exp(i\varphi) & 0 \\ 0 & A \end{pmatrix}, \quad \varphi \in \mathbb{R}, \ A \in \mathrm{U}(n).$$

We see that

$$\mathbb{C}P^n = \mathrm{U}(n+1)/(\mathrm{U}(1) \times \mathrm{U}(n)).$$

When normalizing the direction vectors to make them of unit length, we in fact consider the unit sphere in $\mathbb{C}^{n+1}$. The unit vectors $\xi_1$ and $\xi_2$ specify the same point in $\mathbb{C}P^n$ if and only if they are proportional, $\xi_1 = \exp(i\varphi)\xi_2$. We see that the group $\mathrm{U}(1) = \{\exp(i\varphi)\}$ acts on the sphere $S^{2n+1}$ by multiplications: $\xi \to \exp(i\varphi)\xi$. This action is not transitive, and its space of orbits is the smooth manifold $\mathbb{C}P^n$:

$$\mathbb{C}P^n = S^{2n+1}/\mathrm{U}(1).$$

Since $\mathbb{C}P^1 = S^2$ and $\mathrm{U}(1) = S^1$, we have a remarkable projection

$$S^3 \xrightarrow{S^1} S^2 = S^3/S^1$$

of the three-dimensional sphere onto the two-dimensional sphere, where each layer (the inverse image of a point) is a circle. This mapping is called the *Hopf fibration.*

4. The *Stiefel manifold* $V_{n,k}$ consists of all orthonormal collections $x = (f_1, \dots, f_k)$ of $k$ vectors in $\mathbb{R}^n$. Each of these collections is an orthonormal basis in a $k$-dimensional subspace. The group $\mathrm{O}(n)$ acts on $V_{n,k}$ linearly:

$$(f_1, \dots, f_k) \to (Af_1, \dots, Af_k), \quad A \in \mathrm{O}(n),$$

and it is easily seen that this action is transitive. Let $e_1, \dots, e_n$ be a fixed orthonormal basis in $\mathbb{R}^n$. Consider the point $x = (e_1, \dots, e_k) \in V_{n,k}$. It is

obvious that its isotropy group consists of all matrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}, \quad A \in \mathrm{O}(n-k), \text{ 1 is the unity of the group } \mathrm{O}(k).$$

Therefore

$$V_{n,k} = \mathrm{O}(n)/\mathrm{O}(n-k), \quad \dim V_{n,k} = nk - \frac{k(k+1)}{2}.$$

For each point $x = (f_1, \ldots, f_k) \in V_{n,k}$, let us expand the vectors $f_i$ in the given basis $e_1, \ldots, e_n$:

$$f_i = \sum_{j=1}^{n} x_{ji} e_j, \quad i = 1, \ldots, k.$$

Then we obtain the mapping

$$V_{n,k} \to \mathbb{R}^{nk}\colon (f_1, \ldots, f_k) \to (x_{11}, \ldots, x_{n1}, \ldots, x_{1k}, \ldots, x_{nk}). \tag{6.24}$$

The images of points of $V_{n,k}$ satisfy the orthonormality equations $\langle f_i, f_j \rangle = \delta_{ij}$, which can be rewritten in the coordinate form as

$$\sum_{m=1}^{n} x_{mi} x_{mj} = \delta_{ij}, \quad i, j = 1, \ldots, k, \ i \le j. \tag{6.25}$$

**Lemma 6.6.** *The Stiefel manifold is a regular surface in $\mathbb{R}^{nk}$ specified by equations* (6.25).

**Proof.** It is obvious that each point of the surface (6.25) corresponds to a point of the Stiefel manifold under the mapping (6.24). Since the Stiefel manifold is homogeneous, it suffices to verify regularity of this surface at the point $x_{ij} = \delta_{ij}$. The system (6.25) consists of $\frac{k(k+1)}{2}$ equations, and the tangent space to the surface at the point $x_{ij} = \delta_{ij}$ consists of the vectors $\xi_{ij} = \dot{x}_{ij}$ satisfying the equation

$$0 = \frac{d}{dt}\left(\sum_{m=1}^{n} x_{mi} x_{mj}\right)\Bigg|_{t=0} = \sum_{m=1}^{n} (\xi_{mi}\delta_{mj} + \delta_{mi}\xi_{mj}) = \xi_{ji} + \xi_{ij} = 0.$$

This condition specifies a subspace of dimension $nk - \frac{k(k+1)}{2}$ in the tangent space to $\mathbb{R}^{nk}$. Therefore, the surface given by equations (6.25) is regular. $\square$

Many manifolds considered above are Stiefel manifolds:

$$V_{n,n} = \mathrm{O}(n), \qquad V_{n,n-1} = \mathrm{SO}(n), \qquad V_{n,1} = S^{n-1}.$$

If $k < n$, then the group $\mathrm{SO}(n)$ acts transitively on a Stiefel manifold, and we have

$$V_{n,k} = \mathrm{SO}(n)/\mathrm{SO}(n-k).$$

5. The *Grassmann manifold* (or the *Grassmannian*) $G_{n,k}$ parametrizes all $k$-dimensional linear subspaces in $\mathbb{R}^n$. On $G_{n,k}$ the group $\mathrm{O}(n)$ acts transitively, and the stationary group of the subspace $x^{n-k+1} = \cdots = x^n = 0$ consists of all matrices of the form

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \quad A \in \mathrm{O}(k),\ B \in \mathrm{O}(n-k).$$

Hence

$$G_{n,k} = \mathrm{O}(n)/(\mathrm{O}(k) \times \mathrm{O}(n-k)), \quad \dim G_{n,k} = nk - k^2.$$

On the Stiefel space $V_{n,k}$, define the action of the group $\mathrm{O}(k)$ transforming the frames $(f_1, \dots, f_k) \in V_{n,k}$ by the rule

$$(f_1, \dots, f_k) \to (Af_1, \dots, Af_k) = (a_1^i f_i, \dots, a_n^i f_n),$$

where $A = (a_j^i)$ is an orthogonal matrix, $A \in \mathrm{O}(k)$. This action preserves the subspace spanned by the vectors $f_1, \dots, f_k$. The space of orbits of this action is the Grassmann manifold $G_{n,k}$, and we obtain the projection

$$V_{n,k} \xrightarrow{\mathrm{O}(k)} G_{n,k}.$$

To each $k$-dimensional subspace in $\mathbb{R}^n$ uniquely corresponds its orthogonal complement, and this correspondence determines a diffeomorphism

$$G_{n,k} = G_{n,n-k}.$$

For $k = 1$ we, obviously, have

$$G_{n,1} = \mathbb{R}P^{n-1}.$$

If we consider the oriented $k$-dimensional subspaces in $\mathbb{R}^n$, we will obtain the Grassmann manifold of oriented subspaces. It is denoted by $\tilde{G}_{n,k}$, and we have

$$\tilde{G}_{n,k} = \mathrm{SO}(n)/(\mathrm{SO}(k) \times \mathrm{SO}(n-k)).$$

There is the mapping of "forgetting orientation"

$$\tilde{G}_{n,k} \to G_{n,k},$$

under which every point of $G_{n,k}$ has exactly two inverse images. For $n = 1$, $k = 2$ we have the mapping that glues two antipodal points of the two-dimensional sphere,

$$\tilde{G}_{3,2} = S^2 \to G_{3,2} = S^2/\mathbb{Z}_2 = \mathbb{R}P^2.$$

The general description of the spaces $\tilde{G}_{n,2}$ is given by the following theorem.

**Theorem 6.16.** *The manifold $\tilde{G}_{n,2}$ is complex diffeomorphic to the quadric defined in $\mathbb{C}P^{n-1}$ by the equation*

$$z_1^2 + \cdots + z_n^2 = 0.$$

**Proof.** In each two-dimensional subspace, we take a positively oriented orthonormal basis $\eta_1 = \sum_k a_k e_k$, $\eta_2 = \sum_k b_k e_k$, where $e_1, \dots, e_n$ is an orthonormal basis in $\mathbb{R}^n$. We put $z_1 = a_1 + ib_1, \dots, z_n = a_n + ib_n$. The orthonormality condition for the basis $\eta_1, \eta_2$ can be written as

$$z_1^2 + \dots + z_n^2 = |\eta_1|^2 - |\eta_2|^2 + i\langle \eta_1, \eta_2 \rangle = 0.$$

To each nonzero solution $(z_1 : \dots : z_n)$ of this equation corresponds the oriented plane in $\mathbb{R}^n$ spanned by the vectors $\eta_1, \eta_2$, and to each plane corresponds a solution of this equation determined up to multiplication by a nonzero constant $\lambda \in \mathbb{C}$. The proof is completed. $\square$

6. The *complex Stiefel manifold* $V_{n,k}^{\mathbb{C}}$ consists of all orthonormal systems of $k$ vectors in $\mathbb{C}^n$ with Hermitian scalar product. It can be proved similarly to the real case that

$$V_{n,k}^{\mathbb{C}} = \mathrm{U}(n)/\mathrm{U}(n-k), \quad \dim V_{n,k}^{\mathbb{C}} = 2nk - k^2,$$

and for $k < n$ we have

$$V_{n,k}^{\mathbb{C}} = \mathrm{SU}(n)/\mathrm{SU}(n-k).$$

The simplest examples of such manifolds are

$$V_{n,n}^{\mathbb{C}} = \mathrm{U}(n), \qquad V_{n,1}^{\mathbb{C}} = S^{2n-1}.$$

7. The *complex Grassmann manifold* $G_{n,k}^{\mathbb{C}}$ is formed by all $k$-dimensional linear subspaces in $\mathbb{C}^n$. Again, similarly to the real case, it can be easily shown that

$$G_{n,k}^{\mathbb{C}} = G_{n,n-k}^{\mathbb{C}} = \mathrm{U}(n)/(\mathrm{U}(k) \times \mathrm{U}(n-k)), \quad \dim G_{n,k}^{\mathbb{C}} = 2(n-k)k,$$

and by the definition of the complex projective space,

$$G_{n,1}^{\mathbb{C}} = \mathbb{C}P^{n-1}.$$

8. The Stiefel and Grassmann manifolds can also be defined for vector spaces $\mathbb{H}^n$, where $\mathbb{H}$ is the quaternion algebra. In this case

$$V_{n,k}^{\mathbb{H}} = \mathrm{Sp}(n)/\mathrm{Sp}(n-k), \quad V_{n,n}^{\mathbb{H}} = \mathrm{Sp}(n), \quad G_{n,k}^{\mathbb{H}} = \mathrm{Sp}(n)/(\mathrm{Sp}(k) \times \mathrm{Sp}(n-k)).$$

The manifold $G_{n,1}^{\mathbb{H}}$ is called the *quaternion projective space* $\mathbb{H}P^{n-1}$.

9. Let the vector space $\mathbb{C}^n = \mathbb{R}^{2n}$ be equipped with a Hermitian positive definite scalar product $\langle u, v \rangle$. Take an orthonormal basis $e_1, \dots, e_n$ and denote by $z^k = x^k + iy^k$, $x^k, y^k \in \mathbb{R}$, $k = 1, \dots, n$, the corresponding coordinates. The scalar product of two vectors has the form

$$\langle z, z' \rangle = \sum_{k=1}^{n} (x^k x'^k + y^k y'^k) - i \sum_{k=1}^{n} (x^k y'^k - x'^k y^k),$$

i.e., its real and imaginary parts specify the Euclidean and symplectic scalar products on $\mathbb{R}^{2n}$, respectively.

Recall that an $n$-dimensional subspace in $\mathbb{R}^{2n}$ is called a Lagrange plane if the restriction of the symplectic product to it is identically zero. Take vectors $e'_1, \dots, e'_n$ in the Lagrange plane (relative to the scalar product $\mathrm{Im}\langle\cdot,\cdot\rangle$) such that $\mathrm{Re}\langle e'_j, e'_k\rangle = \delta_{jk}$. Since the plane is Lagrangian, this implies that $\langle e'_j, e'_k\rangle = \delta_{jk}$ and the linear transformation $A\colon e_k \to e'_k$, $k = 1, \dots, n$, is unitary: $A \in \mathrm{U}(n)$. Therefore, any Lagrange plane in $\mathbb{C}^n$ can be obtained from the plane $y^1 = \cdots = y^n = 0$ by a unitary transformation. Obviously, this transformation is defined up to orthogonal transformations of the plane $y^1 = \cdots = y^n = 0$. Thus we conclude that the Lagrange planes in $\mathbb{C}^n$ are uniquely parametrized by the points of the homogeneous space

$$\Lambda(n) = \mathrm{U}(n)/\mathrm{O}(n),$$

which is called a *Lagrangian Grassmann manifold* or *Lagrangian Grassmannian.*

**6.1.6. Complex Lie groups.** A Lie group $G$ is called a *complex Lie group* if it is a complex manifold and the group operations, i.e., multiplication $(g, h) \to gh$ and inversion $g \to g^{-1}$, are given by complex-analytic functions.

Some examples of complex Lie groups were given in Section 3.

1. The group $\mathrm{GL}(n, \mathbb{C})$, which is the domain in $\mathbb{C}^{n^2} = \mathrm{M}(n, \mathbb{C})$ specified by the condition $\det A \neq 0$.

2. The group $\mathrm{SL}(n, \mathbb{C})$ is the complex manifold in $\mathbb{C}^{n^2}$ defined by the equation $\det A = 1$.

3. The group $\mathrm{O}(n, \mathbb{C})$ is defined in $\mathrm{M}(n, \mathbb{C}) = \mathbb{C}^{n^2}$ by the equation

$$A^\top A = 1.$$

It is formed by complex matrices that preserve the scalar product in $\mathbb{C}^n$,

$$\langle \xi, \eta\rangle = \xi^1\eta^1 + \cdots + \xi^n\eta^n, \quad \xi = (\xi^1, \dots, \xi^n),\ \eta = (\eta^1, \dots, \eta^n) \in \mathbb{C}^n.$$

By the basis change

$$e_1, \dots, e_k \to e_1, \dots, e_k, \qquad e_{k+1}, \dots, e_n \to ie_{k+1}, \dots, ie_n$$

in $\mathbb{C}^n$ this scalar product is reduced to the form

$$\langle \xi, \eta\rangle = \xi^1\eta^1 + \cdots + \xi^k\eta^k - \xi^{k+1}\eta^{k+1} - \cdots - \xi^n\eta^n.$$

Hence the group $\mathrm{O}(k, n-k, \mathbb{C})$ of complex matrices that preserve this scalar product is isomorphic to $\mathrm{O}(n, \mathbb{C})$ for any $k = 0, \dots, n$.

4. The group $\mathrm{Sp}(n, \mathbb{C})$ consists of complex matrices of order $2n$ preserving the symplectic form in $\mathbb{C}^{2n}$ specified by the matrix

$$G = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{pmatrix}.$$

Arguing as in Section 3 we can prove the following fact.

**Lemma 6.7.** *Each of the complex Lie groups $G = \mathrm{GL}(n, \mathbb{C})$, $\mathrm{SL}(n, \mathbb{C})$, $\mathrm{O}(n, \mathbb{C})$, or $\mathrm{Sp}(n, \mathbb{C})$ contains as a subgroup the real Lie group $H = \mathrm{GL}(n, \mathbb{R})$, $\mathrm{SL}(n, \mathbb{R})$, $\mathrm{O}(n, \mathbb{R})$, or $\mathrm{Sp}(n, \mathbb{R})$, respectively, consisting of real matrices that satisfy the conditions which determine the corresponding group.*

*The real tangent algebra $\mathfrak{h}$ to a subgroup $H$ generates over the field of complex numbers the tangent algebra $\mathfrak{g}$ to the group $G$.*

In the context of this lemma, the complex Lie algebra $\mathfrak{g}$ is called the complexification of the Lie algebra $\mathfrak{h}$ because it is generated by basis vectors with the same commutation relations but considered over the field of complex numbers. This is denoted by $\mathfrak{g} = \mathfrak{h}^{\mathbb{C}} = \mathfrak{h} \otimes \mathbb{C}$.

As a rule, complex Lie groups are not compact. This is stated by the following theorem.

**Theorem 6.17.** *Any compact connected complex Lie group is commutative and diffeomorphic to the complex torus.*

**Proof.** Consider the homomorphism Ad of the Lie group $G$ into the group $\mathrm{GL}(2n, \mathbb{R})$, where $\dim_{\mathbb{C}} G = n$ and $\mathrm{GL}(2n, \mathbb{R})$ is regarded as a matrix group of automorphisms of the tangent Lie algebra $\mathfrak{g}$. It is specified by the formula

$$\mathrm{Ad}(g)\xi = g\xi g^{-1}, \quad \xi \in \mathfrak{g}$$

(see Section 4).

The matrix coefficients $\mathrm{Ad}\, g$ are holomorphic functions on a compact connected Lie group, hence they are constant (by Theorem 5.11). Since the group $G$ is connected and $\mathrm{Ad}\, 1 = 1$, we obtain that $\mathrm{Ad}\, G = 1$. We conclude that $\mathrm{Ad}(g)\xi = \xi$. Let $g = e^{t\eta}$. As $t \to 0$, the equality $\mathrm{Ad}(g)\xi = \xi$ in the limit turns into the equality $[\eta, \xi] = 0$. Therefore, the Lie algebra $\mathfrak{g}$ is commutative and the group itself is diffeomorphic to the quotient group of the group $\mathbb{R}^{2n}$ with addition operation. Thus the group $G$ is complex diffeomorphic to the complex torus $\mathbb{C}^n/\Lambda$. This completes the proof. $\square$

**6.1.7. Classification of Lie algebras.** Here we present without proof important results on classification of Lie algebras. We will assume that the Lie algebras are defined either over $\mathbb{R}$ or over $\mathbb{C}$.

The Killing form

$$\langle X, Y\rangle = -\operatorname{Tr}(\operatorname{ad}(X)\operatorname{ad}(Y))$$

plays an important role in classification of Lie algebras.

A Lie algebra $\mathfrak{g}$ is said to be *nilpotent* of degree $N$ if

$$[\xi_1, [\xi_2, \dots [\xi_N, \xi_{N+1}] \dots]] = 0$$

for any $\xi_1, \dots, \xi_{N+1} \in \mathfrak{g}$, or, equivalently,

$$\operatorname{ad}(\xi_1) \cdots \operatorname{ad}(\xi_N) = 0$$

for all $\xi_1, \dots, \xi_N \in \mathfrak{g}$. The following Engel theorem establishes a relationship between these concepts.

**Theorem 6.18.** *A Lie algebra is nilpotent if and only if its Killing form is identically zero.*

A subalgebra $\mathfrak{h}$ of a Lie algebra $\mathfrak{g}$ is called an *ideal* if $[\mathfrak{g}, \mathfrak{h}] \subset \mathfrak{h}$. The simplest example of an ideal is the *commutator subalgebra*, denoted by $[\mathfrak{g}, \mathfrak{g}]$, which is the subalgebra linearly generated by the elements of the form $[\xi, \eta]$, where $\xi, \eta \in \mathfrak{g}$.

A Lie algebra $\mathfrak{g}$ is said to be *solvable* if the sequence $D^i\mathfrak{g} = [D^{i-1}\mathfrak{g}, D^{i-1}\mathfrak{g}]$ vanishes after finitely many steps:

$$0 = D^N\mathfrak{g} = [D^{N-1}\mathfrak{g}, D^{N-1}\mathfrak{g}] \subset D^{N-1}\mathfrak{g} \subset \cdots \subset D^1\mathfrak{g} = [\mathfrak{g}, \mathfrak{g}] \subset \mathfrak{g}.$$

It is easily seen that every nilpotent Lie algebra is solvable, but the converse is not true, as the following example shows.

EXAMPLE. Consider the three-dimensional Lie algebra formed by all matrices

$$\begin{pmatrix} e^z & 0 & x \\ 0 & e^{-z} & y \\ 0 & 0 & 1 \end{pmatrix}, \quad x, y, z \in \mathbb{R},$$

with usual matrix multiplication. In these coordinates the multiplication becomes

$$(x, y, z) \cdot (x', y', z') = (e^z x' + x, e^{-z} y' + y, z + z').$$

The Lie algebra of this group is generated by the matrices

$$e_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad e_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which satisfy the commutation relations

$$[e_1, e_2] = 0, \qquad [e_3, e_1] = e_1, \qquad [e_3, e_2] = -e_2.$$

It can be easily verified that this Lie algebra is solvable but not nilpotent (e.g., $[e_3, [e_3, \ldots [e_3, e_1] \ldots]] = e_1$).

The following theorem gives a condition for solvability of a Lie algebra in terms of the Killing form.

**Theorem 6.19.** *A Lie algebra* $\mathfrak{g}$ *is solvable if and only if*

$$\langle X, Y \rangle = 0 \quad \textit{for all } X \in \mathfrak{g},\ Y \in [\mathfrak{g}, \mathfrak{g}].$$

This theorem is a consequence of the following Lie theorem.

**Theorem 6.20.** *A Lie algebra* $\mathfrak{g}$ *over the field of complex numbers* $\mathbb{C}$ *is solvable if and only if there is a basis in* $\mathfrak{g}$ *in which all operators* $\operatorname{ad} X : \mathfrak{g} \to \mathfrak{g}$ *are specified by upper triangular matrices.*

A Lie group is said to be *nilpotent* or *solvable* if its Lie algebra is nilpotent or solvable, respectively.

Let $H$ be a discrete subgroup of a group $G$ and let the manifold $G/H$ be compact. Then such a manifold is called a *nilmanifold* if the group $G$ is nilpotent, and a *solvmanifold* if the group $G$ is solvable.

We see that the Killing forms on solvable and, in particular, nilpotent Lie algebras are degenerate.

If the Killing form on a Lie algebra $\mathfrak{g}$ is nondegenerate, then the algebra $\mathfrak{g}$ is said to be *semisimple*. This is equivalent to the property that $\mathfrak{g}$ does not contain nonzero commutative ideals (Cartan's criterion).

A Lie algebra $\mathfrak{g}$ is said to be *simple* if it does not contain nontrivial ideals different from zero and $\mathfrak{g}$. If a semisimple Lie algebra contains an ideal $\mathfrak{k}$, then it decomposes into a direct sum

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{k}^\perp, \quad [\mathfrak{k}, \mathfrak{k}^\perp] = 0,$$

where $\mathfrak{k}^\perp$ is the orthogonal complement of $\mathfrak{k}$ relative to the nondegenerate Killing form (and $\mathfrak{k}^\perp$ is also an ideal). Therefore, any semisimple Lie algebra is a direct sum of simple Lie algebras:

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_k, \qquad [\mathfrak{g}_i, \mathfrak{g}_j] = 0 \quad \text{for } i \neq j,$$

and the classification of semisimple Lie algebras reduces to that of simple Lie algebras.

It turns out that the classification of simple Lie algebras simplifies if we consider algebras over the field of complex numbers $\mathbb{C}$.

Recall that the complexification $\mathfrak{g} \otimes \mathbb{C}$ of a real Lie algebra $\mathfrak{g}$ is the Lie algebra over $\mathbb{C}$ determined by the same basis $e_1, \dots, e_n$ and the same commutation relations $[e_i, e_j] = c_{ij}^k e_k$ as the algebra $\mathfrak{g}$. The complexifications of two nonisomorphic real Lie algebras may be isomorphic as Lie algebras over $\mathbb{C}$, which is demonstrated by the following example.

EXAMPLE: $\mathfrak{sl}(2) \otimes \mathbb{C} = \mathfrak{so}(3) \otimes \mathbb{C}$. Let $e_1, e_2, e_3$ be the basis (6.15) in the algebra $\mathfrak{so}(3)$, and $Y_0, Y_1, Y_2$ the basis (6.17) in the algebra $\mathfrak{sl}(2, \mathbb{R})$. We have the following commutation relations:

$$[e_1, e_2] = e_3, \qquad [e_2, e_3] = e_1, \qquad [e_3, e_1] = e_2,$$
$$[Y_0, Y_1] = -2Y_2, \qquad [Y_0, Y_2] = 2Y_1, \qquad [Y_1, Y_2] = 2Y_0.$$

The isomorphism of the complex algebras $\mathfrak{sl}(2) \otimes \mathbb{C}$ and $\mathfrak{so}(3) \otimes \mathbb{C}$ is given by the mapping

$$\frac{1}{2} Y_0 \to e_1, \qquad \frac{i}{2} Y_1 \to e_2, \qquad -\frac{i}{2} Y_2 \to e_3.$$

This is only a particular example of such an isomorphism: it was proved above that $\mathfrak{sl}(2) = \mathfrak{so}(1, 2)$, and the algebra $\mathfrak{so}(p, q) \otimes \mathbb{C}$ for any $p, q$ is the Lie algebra of a group of matrices in $\mathrm{M}(n, \mathbb{C})$, $n = p + q$, that preserve the scalar product $\langle u, v\rangle = u^1 v^1 + \cdots + u^n v^n$ in $\mathbb{C}^n$. If we restrict this scalar product to different $n$-dimensional real subspaces in $\mathbb{R}^{2n} = \mathbb{C}^n$, we will obtain different Euclidean and pseudo-Euclidean scalar products in $\mathbb{R}^n$. The Lie algebras of the groups of motions related to these scalar products are different and equal to $\mathfrak{so}(p, q)$, whereas their complexifications are isomorphic.

We say that $\mathfrak{g}$ is the *real form* of a Lie algebra $\mathfrak{g}^{\mathbb{C}}$ over the field $\mathbb{C}$ if $\mathfrak{g} \otimes \mathbb{C} = \mathfrak{g}^{\mathbb{C}}$. We see that the algebras $\mathfrak{so}(k, n-k)$ for $k = 0, \dots, n$ are the real forms of the same complex algebra.

The following Killing–Cartan theorem provides a complete classification of simple complex Lie algebras.

**Theorem 6.21.** *Let $\mathfrak{g}$ be a simple Lie algebra over the field of complex numbers.*

*If it is commutative, then $\mathfrak{g} = \mathbb{C}$.*

*If the algebra $\mathfrak{g}$ is noncommutative, then it belongs to the following list of pairwise nonisomorphic algebras:*

1) $A_n = \mathfrak{su}(n+1) \otimes \mathbb{C}$ *for* $n \geq 1$;

2) $B_n = \mathfrak{so}(2n+1) \otimes \mathbb{C}$ *for* $n \geq 2$;

3) $C_n = \mathfrak{sp}(n) \otimes \mathbb{C}$ *for* $n \geq 3$;

4) $D_n = \mathfrak{so}(2n) \otimes \mathbb{C}$ *for* $n \geq 4$;

5) *the special simple Lie algebras* $E_6, E_7, E_8, F_4$, *and* $G_2$.

*The dimensions of the special algebras are*: $\dim E_6 = 78$, $\dim E_7 = 133$, $\dim E_8 = 248$, $\dim F_4 = 52$, *and* $\dim G_2 = 14$.

The subscript in the notation of these algebras has a special meaning: it indicates the dimension of the maximal commutative subalgebra of the algebra $\mathfrak{g}$.

A Lie group is said to be *simple* if its Lie algebra is simple. In a similar way, a Lie group is *semisimple* if its Lie algebra is semisimple.

A real Lie algebra is said to be *compact* if it is the tangent algebra of a compact Lie group. It turns out that:

1) For semisimple Lie algebras, compactness is equivalent to positive definiteness of the Killing form.

2) Any simple complex Lie algebra has exactly one compact real form.

3) Any compact Lie group is locally isomorphic to the direct product of simple compact Lie groups and a torus $T^n$.

**6.1.8. Two-dimensional and three-dimensional Lie algebras.** For small dimensions there is a complete classification of Lie algebras up to isomorphism. We will present this classification for two-dimensional and three-dimensional Lie algebras.

1. Let $\mathfrak{g}$ be a two-dimensional Lie algebra with basis $e_1, e_2$. Then multiplication in this algebra is specified by a single relation

$$[e_1, e_2] = ae_1 + be_2,$$

since $[e_1, e_1] = [e_2, e_2] = 0$ and $[e_2, e_1] = -[e_1, e_2]$ by the definition of Lie algebras.

If $a = b = 0$, then the Lie algebra is commutative.

Suppose that multiplication in the algebra $\mathfrak{g}$ is nontrivial. Without loss of generality, assume that $a \neq 0$ and select the new basis

$$e_1' = -\frac{1}{a}\, e_2, \qquad e_2' = e_1 + \frac{b}{a}\, e_2.$$

Then the commutation relation becomes

$$[e_1', e_2'] = e_2'. \tag{6.26}$$

**Lemma 6.8.** *The real two-dimensional Lie algebra with commutation relation* (6.26) *is isomorphic to the Lie algebra of the group of affine transformations of the real line* $\mathbb{R}$.

**Proof.** The group A(1) of affine transformations of the real line consists of transformations of the form $x \to \lambda x + \mu$.

Consider the imbedding

$$x \to (x, 1)$$

of $\mathbb{R}$ into $\mathbb{R}^2$. The group $\mathrm{A}(1)$ can be realized by the matrices $\left(\begin{smallmatrix}\lambda & \mu\\ 0 & 1\end{smallmatrix}\right)$. Indeed, we have

$$\begin{pmatrix}\lambda & \mu\\ 0 & 1\end{pmatrix}\begin{pmatrix}x\\ 1\end{pmatrix} = \begin{pmatrix}\lambda x + \mu\\ 1\end{pmatrix}.$$

The tangent algebra to this group is generated by the matrices

$$Y_1 = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \qquad Y_2 = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix},$$

which satisfy the relation

$$[Y_1, Y_2] = Y_1 Y_2 - Y_2 Y_1 = Y_2,$$

which proves the lemma. □

This lemma implies the following result.

**Theorem 6.22.** *If $\mathfrak{g}$ is a two-dimensional Lie algebra, then it has a basis $e_1, e_2$ such that one of the following two relations holds:*

$$[e_1, e_2] = 0 \quad \text{or} \quad [e_1, e_2] = e_2.$$

*In the former case the algebra $\mathfrak{g}$ is commutative, while in the latter it is isomorphic to the Lie algebra of the group of affine transformations of the real line.*

2. Let $\mathfrak{g}$ be a three-dimensional Lie algebra with basis $e_1, e_2, e_3$ and commutation relations $[e_i, e_j] = c_{ij}^k e_k$, where $c_{ij}^k = -c_{ji}^k$. The Jacobi identity is equivalent to the system of equations (6.9) on the structural constants $c_{ij}^k$:

$$c_{il}^m c_{jk}^l + c_{jl}^m c_{ki}^l + c_{kl}^m c_{ij}^l = 0.$$

Since the constants $c_{ij}^k$, $i, j, k = 1, 2, 3$, are skew-symmetric in subscripts $i, j$, $c_{ij}^k = -c_{ji}^k$, they can easily be represented in the form

$$c_{ij}^k = \varepsilon_{ijl} t^{lk},$$

where the symbol $\varepsilon_{ijk}$ is defined by (6.20) (and is equal to $\pm 1$ depending on whether $\left(\begin{smallmatrix}1 & 2 & 3\\ i & j & k\end{smallmatrix}\right)$ is an even or odd permutation, and to zero if some of the $i, j, k$ coincide).

Let us single out the symmetric and skew-symmetric parts of the matrix $t^{lk}$:

$$b^{lk} = \frac{1}{2}(t^{lk} + t^{kl}), \qquad d^{lk} = \frac{1}{2}(t^{lk} - t^{kl}),$$

$$t^{lk} = b^{lk} + d^{lk}, \qquad b^{lk} = b^{kl}, \qquad d^{lk} = -d^{kl}.$$

The space of skew-symmetric $3 \times 3$ matrices $X = (x^{lk})$ is three-dimensional and contains the three-dimensional subspace of matrices of the form

$$x^{lk} = \varepsilon^{ijl}(\delta^k_j a_i - \delta^k_i a_j), \quad a = (a_1, a_2, a_3).$$

Therefore, any matrix, including the matrix $d^{lk}$, is representable in this form. Thus we have shown that the structural constants $c^k_{ij}$ of the three-dimensional Lie algebra are representable in the form

$$c^k_{ij} = \varepsilon_{ijl} b^{lk} + \varepsilon_{ijl} d^{lk} = \varepsilon_{ijl} b^{lk} + \delta^k_j a_i - \delta^k_i a_j,$$

where the matrix $b^{lk}$ is skew-symmetric and $a = (a_1, a_2, a_3)$ is a three-dimensional vector.

Now let $e_1, \dots, e_n$ and $\tilde{e}_1, \dots, \tilde{e}_n$ be two bases in the Lie algebra $\mathfrak{g}$ related by the transition formulas

$$\tilde{e}_p = a^i_p e_i, \quad e_i = b^p_i \tilde{e}_p, \quad a^p_i b^j_p = \delta^j_i.$$

Then the corresponding structural constants $c^k_{ij}$ and $\tilde{c}^r_{pq}$ are related by

$$c^k_{ij} = a^k_r b^p_i b^q_j \tilde{c}^r_{pq}. \tag{6.27}$$

This formula is derived by simple calculations:

$$[e_i, e_j] = b^p_i b^q_j [\tilde{e}_p, \tilde{e}_q] = b^p_i b^q_j \tilde{c}^r_{pq} \tilde{e}_r = (b^p_i b^q_j \tilde{c}^r_{pq} a^k_r) e_k = c^k_{ij} e_k.$$

Now let $c^k_{ij}$ be structural constants of a three-dimensional Lie algebra $\mathfrak{g}$. The above lemmas imply that by a change of the basis, the $c^k_{ij}$ can be reduced to the form

$$c^k_{ij} = \varepsilon_{ijl} b^{(l)} \delta^{lk} + \delta^k_j a_i - \delta^k_i a_j,$$

i.e., the matrix $b^{lk}$ becomes diagonal, $b^{lk} = b^{(l)} \delta^{lk}$.

The Jacobi identity (system (6.9)) implies that

$$b^{lk} a_k = 0.$$

Since $b^{lk}$ is a diagonal matrix, we can assume without loss of generality that $a = (a, 0, 0)$. Then the commutation relations become

$$[e_1, e_2] = a e_2 + b^{(3)} e_3, \qquad [e_2, e_3] = b^{(1)} e_1, \qquad [e_3, e_1] = b^{(2)} e_2 - a e_3,$$

where

$$a b^{(1)} = 0.$$

Normalizing the vectors $e_1$, $e_2$, $e_3$ we obtain the following classification of the three-dimensional Lie algebras up to isomorphism, known as the Bianchi classification.

| Type | $a$ | $b^{(1)}$ | $b^{(2)}$ | $b^{(3)}$ | Type | $a$ | $b^{(1)}$ | $b^{(2)}$ | $b^{(3)}$ |
|---|---|---|---|---|---|---|---|---|---|
| [2pt] I | 0 | 0 | 0 | 0 | V | 1 | 0 | 0 | 0 |
| II | 0 | 1 | 0 | 0 | IV | 1 | 0 | 0 | 1 |
| $\mathrm{VII}_0$ | 0 | 1 | 1 | 0 | $\mathrm{VII}_a$, $a > 0$ | $a$ | 0 | 1 | 1 |
| $\mathrm{VI}_0$ | 0 | 1 | $-1$ | 0 | III | 1 | 0 | 1 | $-1$ |
| IX | 0 | 1 | 1 | 1 | $\mathrm{VI}_a$, $0 < \lvert a\rvert < 1$ | $a$ | 0 | 1 | $-1$ |
| VIII | 0 | 1 | 1 | $-1$ | | | | | |

The algebras of type I and III are decomposable into direct sums: the algebra of type I is commutative and the algebra of type III is isomorphic to the direct sum $\mathfrak{a}(1) \oplus \mathbb{R}$.

The other algebras are indecomposable; in particular, the Lie algebra of the Heisenberg group is of type II, $\mathfrak{e}(2)$ is of type $\mathrm{VII}_0$, $\mathfrak{sl}(2,\mathbb{R}) = \mathfrak{so}(1,2)$ is of type VIII, and $\mathfrak{so}(3,\mathbb{R})$ is of type IX.

**6.1.9. Poisson structures.** There are important examples of infinite-dimensional Lie algebras. Some of them, to be discussed now, are specified by Poisson structures on spaces of smooth functions; others, Lie algebras of vector fields, will be considered in the next chapter.

A vector space $L$ is called a *Poisson algebra* if it is a Lie algebra, i.e.,

$$[\xi,\eta] = -[\eta,\xi], \qquad [\xi,[\eta,\zeta]] + [\eta,[\zeta,\xi]] + [\zeta,[\xi,\eta]] = 0,$$

and, in addition, a multiplication operation is defined on this space,

$$\xi,\eta \to \xi\eta \in L,$$

which is commutative, associative, linear in both arguments, and satisfies the *Leibniz identity*

$$[\xi\eta,\zeta] = \xi\,[\eta,\zeta] + \eta\,[\xi,\zeta].$$

The most important example of such an operation is given by the Poisson bracket on the space of smooth functions.

Let $M$ be an $n$-dimensional manifold or a domain in $\mathbb{R}^n$. Denote by $C^\infty(M)$ the space of all smooth functions on $M$. Then $M$ is said to be endowed with a *Poisson structure* if on the space $C^\infty(M)$ there is a bilinear operation

$$f,g \to \{f,g\} \tag{6.28}$$

such that the space $C^\infty(M)$ is a Lie algebra relative this operation, i.e.,

$$\{f,g\} = -\{g,f\}, \qquad \{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} = 0,$$

and the Lie algebra $C^\infty(M)$ is a Poisson algebra relative to ordinary multiplication of functions, i.e., the Leibniz identity

$$\{fg,h\} = f\{g,h\} + g\{f,h\}, \quad f,g,h \in C^\infty(M),$$

holds. In this case the operation (6.28) is called the *Poisson bracket.*

A smooth manifold endowed with a Poisson structure is called a *Poisson manifold.*

Let $f$ be the product of powers of coordinate variables, $f(x^1, \dots, x^n) = (x^1)^{a_1} \cdots (x^n)^{a_n}$. Then it can be written as $f = (x^1)^{a_1} h(x^2, \dots, x^n)$. It follows from the Leibniz identity that

$$\begin{aligned}\{f, g\} &= h\{(x^1)^{a_1}, g\} + (x^1)^{a_1}\{h, g\} = h a_1 (x^1)^{a_1 - 1}\{x^1, g\} + (x^1)^{a_1}\{h, g\} \\ &= \frac{\partial f}{\partial x^1}\{x^1, g\} + (x^1)^{a_1}\{h, g\}.\end{aligned}$$

Applying these arguments to the variables $x^2, \dots, x^n$, we obtain

$$\{f, g\} = \frac{\partial f}{\partial x^i}\{x^i, g\}.$$

This relation can be extended by linearity to all polynomials, and more generally to all real-analytic functions $f$. Thus we have proved the following lemma.

**Lemma 6.9.** *If $f, g$ are real-analytic functions of $x^1, \dots, x^n$ in a domain $M \subset \mathbb{R}^n$, and the Poisson bracket is defined in this domain, then this bracket of $f, g$ is given by the formula*

$$\{f, g\} = \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j} h^{ij}(x), \quad \textit{where} \quad \{x^i, x^j\} = h^{ij}(x) = -h^{ji}(x). \tag{6.29}$$

In what follows we will simply require the Poisson bracket to have the form (6.29). Of course, the quantities $h^{ij}$ must satisfy certain conditions.

**Lemma 6.10.** *Suppose that on the space of smooth functions on $M$ a skew-symmetric bilinear operation is defined, which in local coordinates $x^1, \dots, x^n$ has the form* (6.29).

*Then this operation satisfies the Leibniz identity and in arbitrary local coordinates $y^1, \dots, y^n$ has the form*

$$\{f, g\} = \frac{\partial f}{\partial y^k}\frac{\partial g}{\partial y^l}\{y^k, y^l\}.$$

**Proof.** The proof of the Leibniz identity is simple:

$$\begin{aligned}\{fg, h\} &= \frac{\partial (fg)}{\partial x^i}\frac{\partial h}{\partial x^j}\{x^i, x^j\} = \left(f\frac{\partial g}{\partial x^i} + g\frac{\partial f}{\partial x^i}\right)\frac{\partial h}{\partial x^j}\{x^i, x^j\} \\ &= f\{g, h\} + g\{f, h\}.\end{aligned}$$

If we are given a change of coordinates $y^i = y^i(x^1, \ldots, x^n)$, $i = 1, \ldots, n$, then

$$\{f, g\} = \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j}\{x^i, x^j\} = \frac{\partial f}{\partial y^k}\frac{\partial y^k}{\partial x^i}\frac{\partial g}{\partial y^l}\frac{\partial y^l}{\partial x^j}\{x^i, x^j\}$$
$$= \frac{\partial f}{\partial y^k}\frac{\partial g}{\partial y^l}\Big(\frac{\partial y^k}{\partial x^i}\frac{\partial y^l}{\partial x^j}\{x^i, x^j\}\Big) = \frac{\partial f}{\partial y^k}\frac{\partial g}{\partial y^l}\{y^k, y^l\}.$$

The proof is completed. □

The condition that the bilinear operation (6.29) satisfies the Jacobi identity implies additional restrictions on the functions $h^{ij}(x)$. The following lemma provides a simple criterion for fulfillment of this identity.

**Lemma 6.11.** *An operation* $\{f, g\}$ *defined by* (6.29) *satisfies the identity*

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\}$$
$$= \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j}\frac{\partial h}{\partial x^k}(\{x^i, \{x^j, x^k\}\} + \{x^j, \{x^k, x^i\}\} + \{x^k, \{x^i, x^j\}\});$$

*therefore, the Jacobi identity is fulfilled if and only if it holds for linear functions,*

$$\{x^i, \{x^j, x^k\}\} + \{x^j, \{x^k, x^i\}\} + \{x^k, \{x^i, x^j\}\} = 0. \tag{6.30}$$

*In terms of* $h^{ij}(x) = \{x^i, x^j\}$, *relation* (6.30) *is written as*

$$\frac{\partial h^{jk}}{\partial x^l}h^{il} + \frac{\partial h^{ki}}{\partial x^l}h^{jl} + \frac{\partial h^{ij}}{\partial x^l}h^{kl} = 0. \tag{6.31}$$

**Proof.** We have

$$\{f, \{g, h\}\} = \frac{\partial f}{\partial x^i}\{x^i, \{g, h\}\} = \frac{\partial f}{\partial x^i}\Big\{x^i, \frac{\partial g}{\partial x^j}\frac{\partial h}{\partial x^k}\{x^j, x^k\}\Big\}$$
$$= \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j}\frac{\partial h}{\partial x^k}\{x^i, \{x^j, x^k\}\} + \frac{\partial f}{\partial x^i}\frac{\partial}{\partial x^l}\Big(\frac{\partial g}{\partial x^j}\frac{\partial h}{\partial x^k}\Big)\{x^i, x^l\}\{x^j, x^k\}.$$

We sum this equality with its analogs for $\{g, \{h, f\}\}$ and $\{h, \{f, g\}\}$ and write down the left-hand side of the Jacobi identity for the functions $f, g, h$ in the form

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\}$$
$$= \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j}\frac{\partial h}{\partial x^k}(\{x^i, \{x^j, x^k\}\} + \{x^j, \{x^k, x^i\}\} + \{x^k, \{x^i, x^j\}\}) + Q,$$

where $Q = Q_f + Q_g + Q_h$ is the term containing the second-order derivatives of $f, g, h$. Let us demonstrate, for example, the term $Q_f$ containing

the second-order derivatives of $f$. They arise from the terms $\{h, \{f, g\}\}$ and $\{g, \{h, f\}\}$, and so we have

$$
\begin{aligned}
Q_f &= \Big(\frac{\partial h}{\partial x^i} \frac{\partial^2 f}{\partial x^l \partial x^j} \frac{\partial g}{\partial x^k} + \frac{\partial g}{\partial x^i} \frac{\partial h}{\partial x^j} \frac{\partial^2 f}{\partial x^l \partial x^k}\Big) \{x^i, x^l\}\{x^j, x^k\} \\
&= \Big(\frac{\partial h}{\partial x^i} \frac{\partial g}{\partial x^k} - \frac{\partial h}{\partial x^k} \frac{\partial g}{\partial x^i}\Big) \frac{\partial^2 f}{\partial x^l \partial x^j} \{x^i, x^l\}\{x^j, x^k\}.
\end{aligned}
$$

Of course, summation over repeated indices is assumed in this formula, and since permutations of indices of the type $i \leftrightarrow k$, $j \leftrightarrow l$ reverse the signs of the corresponding terms, their sum is equal to zero, $Q_f = 0$. In a similar way, $Q_g = Q_h = 0$.

It remains to note that

$$
\begin{aligned}
&\{x^i, \{x^j, x^k\}\} + \{x^j, \{x^k, x^i\}\} + \{x^k, \{x^i, x^j\}\} \\
&\quad = \{x^i, h^{jk}\} + \{x^j, h^{ki}\} + \{x^k, h^{ij}\} = \frac{\partial h^{jk}}{\partial x^l} h^{il} + \frac{\partial h^{ki}}{\partial x^l} h^{jl} + \frac{\partial h^{ij}}{\partial x^l} h^{kl}.
\end{aligned}
$$

The proof is completed. □

Now we point out some important examples.

**Theorem 6.23.** *If the matrix $h^{ij}$ is constant in the domain $M \subset \mathbb{R}^m$, then formula* (6.29) *determines the Poisson bracket.*

**Proof.** For a constant matrix $h^{ij}$, equations (6.31) obviously hold. □

Of particular interest is the case where the rank of the constant matrix $h^{ij}$ equals the dimension of the space. Then by Theorem 2.1 the Poisson bracket in suitable coordinates $q^1, \dots, q^n, p_1, \dots, p_n$, $m = 2n$, becomes

$$\{f, g\} = \sum_{i=1}^{n} \Big(\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} - \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i}\Big).$$

In this case the Poisson bracket is said to be determined by the *symplectic structure.*

This is the form in which the Poisson bracket first appeared in analytic mechanics.

The *Lie–Poisson bracket* on the space of smooth functions on $\mathbb{R}^n$ has the form (6.29) with elements of the matrix $h^{ij} = \{x^i, x^j\}$ depending linearly on coordinates,

$$\{f, g\} = \frac{\partial f}{\partial x^i} \frac{\partial g}{\partial x^j} c_k^{ij} x^k, \tag{6.32}$$

where

$$\{x^i, x^j\} = h^{ij}(x) = c_k^{ij} x^k.$$

**Theorem 6.24.** *Formula (6.32) determines the Poisson bracket if and only if the quantities $c_k^{ij}$ are the structural constants of a Lie algebra, i.e., there is a Lie algebra $\mathfrak{g}$ with basis $e^1, \ldots, e^n$ such that*

$$[e^i, e^j] = c_k^{ij} e^k.$$

**Proof.** We have

$$\{x^i, \{x^j, x^k\}\} + \{x^j, \{x^k, x^i\}\} + \{x^k, \{x^i, x^j\}\} = (c_m^{il} c_l^{jk} + c_m^{jl} c_l^{ki} + c_m^{kl} c_l^{ij}) x^m,$$

and in view of equation (6.9) the condition that this expression vanishes is equivalent to the fact that the $c_k^{ij}$ are the structural constants of some Lie algebra. □

Note that if the matrix $h^{ij}$ is nonsingular, then the equations (6.31) are linear in terms of the inverse matrix. Namely, introduce the matrix $h_{kl}$ uniquely determined by the equations

$$h_{ik} h^{kj} = \delta_i^j, \quad i, j = 1, \ldots, n.$$

Since the matrix $h^{ij}$ is skew-symmetric, $h_{kl}$ is also skew-symmetric, i.e., $h_{kl} = -h_{lk}$, and we have

$$h^{ik} h_{kj} = -h^{ik} h_{jk} = \delta_i^j, \quad i, j = 1, \ldots, n.$$

**Theorem 6.25.** *A skew-symmetric matrix $h^{ij}$ nonsingular in a domain $U \subset \mathbb{R}^n$ determines the Poisson bracket in this domain if and only if the inverse matrix $h_{kl}$ satisfies the "Maxwell equations"*

$$\frac{\partial h_{rs}}{\partial x^p} + \frac{\partial h_{sp}}{\partial x^r} + \frac{\partial h_{pr}}{\partial x^s} = 0, \quad p, r, s = 1, \ldots, n.$$

**Proof.** Note first of all that

$$\frac{\partial (h_{ik} h^{kj})}{\partial x^l} = h_{ik} \frac{\partial h^{kj}}{\partial x^l} + \frac{\partial h_{ik}}{\partial x^l} h^{kj} = 0.$$

By Lemma 6.11, the matrix $h^{ij}$ determines the Poisson bracket if and only if

$$\frac{\partial h^{jk}}{\partial x^l} h^{il} + \frac{\partial h^{ki}}{\partial x^l} h^{jl} + \frac{\partial h^{ij}}{\partial x^l} h^{kl} = 0, \quad i, j, k = 1, \ldots, n$$

(recall that we assume summation over repeated subscripts and superscripts, in this case over $l = 1, \ldots, n$). Multiply these equations by $h_{pi} h_{rj} h_{sk}$ and sum over repeated indices $i, j, k$:

$$h_{pi} h_{rj} h_{sk} \Big( \frac{\partial h^{jk}}{\partial x^l} h^{il} + \frac{\partial h^{ki}}{\partial x^l} h^{jl} + \frac{\partial h^{ij}}{\partial x^l} h^{kl} \Big) = 0. \tag{6.33}$$

Consider, for example, one of the terms,

$$h_{pi} h_{rj} h_{sk} \frac{\partial h^{jk}}{\partial x^l} h^{il} = \delta_p^l h_{rj} h_{sk} \frac{\partial h^{jk}}{\partial x^l} = h_{rj} h_{sk} \frac{\partial h^{jk}}{\partial x^p}.$$

Taking into account that $h_{rj}\frac{\partial h^{jk}}{\partial x^p} = -\frac{\partial h_{rj}}{\partial x^p}h^{jk}$, rewrite this term as

$$-h_{sk}\frac{\partial h_{rj}}{\partial x^p}h^{jk} = \delta^j_s\frac{\partial h_{rj}}{\partial x^p} = \frac{\partial h_{rs}}{\partial x^p}.$$

By applying the same arguments to the other terms, equations (6.33) become

$$\frac{\partial h_{rs}}{\partial x^p} + \frac{\partial h_{sp}}{\partial x^r} + \frac{\partial h_{pr}}{\partial x^s} = 0.$$

This completes the proof. □

**6.1.10. Graded algebras and Lie superalgebras.** The general definition of an algebra is as follows: an *algebra* $L$ is a vector space over a given field $F$ with a multiplication operation

$$L \times L \to L\colon (a, b) \to ab,$$

which is bilinear in both factors, i.e.,

$$(x_1a_1 + x_2a_2)b = x_1(a_1b) + x_2(a_2b),$$
$$a(x_1b_1 + x_2b_2) = x_1(ab_1) + x_2(ab_2)$$

for all $a_1, a_2, b_1, b_2 \in L$ and $x_1, x_2 \in F$.

The algebra is *commutative* if

$$ab = ba$$

for all $a, b \in L$, and *associative* if

$$(ab)c = a(bc)$$

for all $a, b, c \in L$.

An important example of a nonassociative and a noncommutative algebra is provided by a Lie algebra with commutation operation $[\xi, \eta]$.

An algebra $L$ is called *graded* if it decomposes into a family of subspaces

$$L = \bigoplus_{\alpha \in G} L_\alpha$$

parametrized by elements of a commutative semigroup $G$, such that

$$L_\alpha \cdot L_\beta \subset L_{\alpha+\beta}$$

for this family. The elements lying in the same subspace $L_\alpha$ are called homogeneous.

The most important are the cases when $G$ is the group of integers $\mathbb{Z}$, the semigroup of nonnegative integers $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$, or the group $\mathbb{Z}_2$. However, the $\mathbb{Z}^+$-graded algebras may be regarded as a particular case of $\mathbb{Z}$-graded algebras with $L_i = 0$ for $i < 0$.

EXAMPLES. 1. With every finite-dimensional Lie algebra $\mathfrak{g}$ one can associate the graded Lie algebra $\mathfrak{g}[t,t^{-1}]$ of Laurent polynomials in $t$ with coefficients in $\mathfrak{g}$:

$$\xi = \sum_{-d\le i\le d} \xi_i t^i, \quad \xi_i \in \mathfrak{g}, \quad i = 0,1,2,\dots, \quad 0 \le d = d(\xi) < \infty.$$

The commutator is defined by termwise commutation of coefficients:

$$\left[\sum_i \xi_i t^i, \sum_j \eta_j t^j\right] = \sum_k \Big(\sum_{i+j=k} [\xi_i, \eta_j]\Big) t^k. \tag{6.34}$$

The coefficient of each power of $t$ is a finite sum of commutators of the form $[\xi_i, \eta_j]$ and hence is well defined. Such algebras are called *affine Lie algebras.*

They may be extended to the algebra $\mathfrak{g}((t))$ of Laurent series with coefficients in $\mathfrak{g}$:

$$\xi = \sum_{i\ge i_0>-\infty} \xi_i t^i.$$

In this case the commutator on the Lie algebra $\mathfrak{g}((t))$ is again well defined by formula (6.34) because the coefficients of powers of $t$ in the right-hand side again are defined by finite sums of commutators.

This algebra contains the subalgebra $\mathfrak{g}[[t]]$ consisting of formal series in nonnegative powers of $t$ with coefficients in $\mathfrak{g}$.

2. An algebra $L$ over a normed field $F$ (e.g., over $\mathbb{R}$ or $\mathbb{C}$) is called *normed* if there is a nonnegative function

$$L \to \mathbb{R}\colon \eta \to \|\eta\|$$

satisfying the following conditions:

1) $\|\lambda\eta\| = |\lambda|\|\eta\|$ for all $\eta \in L$ and $\lambda \in F$;
2) $\|\eta_1 + \eta_2\| \le \|\eta_1\| + \|\eta_2\|$ for all $\eta_1, \eta_2 \in L$;
3) $\|\eta\| = 0$ if and only if $\eta = 0$;
4) $\|\eta \cdot \zeta\| \le \|\eta\|\|\zeta\|$ for all $\eta, \zeta \in L$.

If $\mathfrak{g}$ is a normed Lie algebra, then the algebra $\mathfrak{g}((t))$ may be extended to the *algebra of currents* (*current algebra*) $\tilde{\mathfrak{g}}$, consisting of series

$$\xi = \sum_{i\in\mathbb{Z}} \xi_i t^i, \quad \xi_i \in \mathfrak{g},$$

with rapidly decreasing norms of coefficients $\|\xi_i\|$ as $|i| \to \infty$. In this case the sums $\sum_{i+j=k} \|[\xi_i, \eta_j]\|$ converge for all such series $\xi$ and $\eta$, hence the commutators of the elements of this current algebra are well defined by formula (6.34).

For the one-dimensional commutative Lie algebra $\mathfrak{g} = \mathbb{C}$, the current algebra coincides with the algebra $L_2(S^1)$ of $2\pi$-periodic complex-valued

square-integrable functions $f(x)$, i.e., such that $\int_0^{2\pi} |f(x)|^2\,dx < \infty$. The isomorphism is established by the mapping $t \to e^{2\pi i x}$, which takes a series in $\tilde{\mathfrak{g}}$ into the Fourier series of the corresponding function in $L_2(S^1)$.

The algebras in Examples 1 and 2 are $\mathbb{Z}$-graded in powers of $t$.

3. A Lie algebra $\mathfrak{g}$ and its subalgebra $\mathfrak{k}$ form a *symmetric pair* $(\mathfrak{g}, \mathfrak{k})$ if there exists a $\mathbb{Z}_2$-grading of $\mathfrak{g}$ such that $\mathfrak{g}_0 = \mathfrak{k}$:

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1, \quad \mathfrak{g}_0 = \mathfrak{k}, \quad [\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subset \mathfrak{g}_{(\alpha+\beta) \bmod 2}.$$

This condition appeared first in the theory of symmetric spaces. It implies that the linear involution

$$\sigma(\xi) = \begin{cases} \xi & \text{for } \xi \in \mathfrak{g}_0, \\ -\xi & \text{for } \xi \in \mathfrak{g}_1 \end{cases}$$

is an automorphism (an isomorphism onto itself) of the Lie algebra $\mathfrak{g}$.

In the 1940s, a generalization of Lie algebras appeared in algebraic topology, namely, Lie superalgebras; later, they proved useful in quantum field theory. By definition, a *Lie superalgebra* $\mathfrak{g}$ is a $\mathbb{Z}_2$-graded algebra

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1$$

with supercommutation operation satisfying the following graded analogs of skew-symmetry and the Jacobi identity:

$$[\xi, \eta] = -(-1)^{\alpha\beta}[\eta, \xi] \quad \text{for} \quad \xi \in \mathfrak{g}_\alpha,\ \eta \in \mathfrak{g}_\beta,$$
$$[\xi, [\eta, \zeta]] = [[\xi, \eta], \zeta] + (-1)^{\alpha\beta}[\eta, [\xi, \zeta]] \quad \text{for} \quad \xi \in \mathfrak{g}_\alpha,\ \eta \in \mathfrak{g}_\beta.$$

The subspace $\mathfrak{g}_0$ of a Lie superalgebra is closed with respect to the supercommutation operation, i.e., $[\mathfrak{g}_0, \mathfrak{g}_0] \subset \mathfrak{g}_0$, and it is easily seen to be the ordinary Lie algebra relative to this operation.

EXAMPLE. Let $L$ be a $\mathbb{Z}_2$-graded algebra with associative multiplication $(\xi, \eta) \to \xi \cdot \eta$. Define a supercommutator $[\,,\,]\colon L \times L \to L$ on $L$. By definition, it is linear in both factors, hence it suffices to define it for homogeneous elements. Set

$$[\xi, \eta] = \xi \cdot \eta - (-1)^{\alpha\beta}\eta \cdot \xi \quad \text{for} \quad \xi \in L_\alpha,\ \eta \in L_\beta. \tag{6.35}$$

It can be easily verified that the algebra $L$ with supercommutation operation is a Lie superalgebra.

An important class of graded algebras is formed by *graded-commutative* (or *skew-commutative*, or *supercommutative*) *algebras*. These are $\mathbb{Z}$-graded algebras $L$ with multiplication that satisfies the identity

$$\xi \cdot \eta = (-1)^{\alpha\beta}\eta \cdot \xi \quad \text{for} \quad \xi \in L_\alpha,\ \eta \in L_\beta,\ \alpha, \beta \in \mathbb{Z}.$$

On every $\mathbb{Z}$-graded algebra there is a natural $\mathbb{Z}_2$-grading:

$$L = \tilde{L}_0 \oplus \tilde{L}_1, \quad \tilde{L}_0 = \bigoplus_k L_{2k}, \quad \tilde{L}_1 = \bigoplus_k L_{2k+1}.$$

The elements of $\tilde{L}_0$ are called even, and those of $\tilde{L}_1$, odd. The supercommutator (6.35) on $L$ is trivial, $[\xi, \eta] = 0$ for all $\xi, \eta \in L$. As usual, Lie superalgebras with trivial commutator are called commutative.

EXAMPLE. Let $V$ be a vector space with basis $e_1, \dots, e_n$. Consider the vector space $\Lambda V$ of all linear combinations of the form

$$e_{i_1} \wedge \cdots \wedge e_{i_k}.$$

Set

$$e_{i_1} \wedge \cdots \wedge e_{i_k} = \operatorname{sgn}(\sigma) e_{\sigma(1)} \wedge \cdots \wedge e_{\sigma(k)} \tag{6.36}$$

and

$$e_{i_1} \wedge \cdots \wedge e_{i_k} = 0 \quad \text{if} \quad i_l = i_m \quad \text{for} \quad l \neq m. \tag{6.37}$$

No other relations are assumed. Define the bilinear multiplication operation on $\Lambda V$ specifying it for the generators by the rule

$$(e_{i_1} \wedge \cdots \wedge e_{i_k}) \wedge (e_{i_1} \wedge \cdots \wedge e_{i_l}) = e_{i_1} \wedge \cdots \wedge e_{i_k} \wedge e_{i_1} \wedge \cdots \wedge e_{i_l}. \tag{6.38}$$

Relations (6.36) and (6.37) imply that the space $\Lambda V$ is finite-dimensional with basis formed by the elements

$$e_{i_1} \wedge \cdots \wedge e_{i_k}, \quad i_1 < \cdots < i_k.$$

On the algebra $\Lambda V$ there is a natural grading

$$e_{i_1} \wedge \cdots \wedge e_{i_k} \in \Lambda^k V.$$

Relations (6.36) and (6.38) immediately imply that the algebra $\Lambda V$ is graded-commutative. It suffices to verify this statement for the generators. Since

$$\operatorname{sgn} \begin{pmatrix} 1 & & k & k+1 & & k+l \\ k+1 & & k+l & 1 & & k \end{pmatrix} = (-1)^{kl},$$

we obtain

$$e_{i_1} \wedge \cdots \wedge e_{i_k} \wedge e_{i_1} \wedge \cdots \wedge e_{i_l} = (-1)^{kl} e_{i_1} \wedge \cdots \wedge e_{i_l} \wedge e_{i_1} \wedge \cdots \wedge e_{i_k}.$$

The graded-commutative algebra $\Lambda V$ is called the *exterior algebra* or *Grassmann algebra* of the space $V$.

If the space $V$ is a Lie algebra, then one can define in a natural way the structure of a Lie superalgebra on its exterior algebra, where the multiplication is determined by the Schouten bracket (see 7.3.3).

## 6.2. Crystallographic groups and their generalizations

**6.2.1. Crystallographic groups in Euclidean spaces.** In physics, the concept of infinitely extended solid body, adopted before the appearance of quantum mechanics, was that the body is formed by atoms that take rigid positions in space, cannot be infinitely close to each other, and at the same time fill out the entire space.

Mathematically this means that the atoms are considered to be point-particles that form a lattice in space.

This notion is defined as follows: a set $\Gamma$ of points in space is called a *lattice* if there are positive constants $R > r > 0$ such that the distance between any two points of $\Gamma$ is at least $r$ and each ball of radius $R$ contains at least one point of $\Gamma$.

Recall that if $G$ is a subgroup of the group of motions of the space, then the *orbit* $Gx$ of a point $x$ is the set of points obtained from $x$ by the transformations in $G$.

A *crystal* is a regular lattice in space. Namely, it is required that the lattice be invariant with respect to the action of some group $G$ formed by motions of the space, and the entire lattice be a union of finitely many orbits of this action:

$$\Gamma = \bigcup_{i=1}^{k} Gx_i.$$

Such a group $G$ is called a *crystallographic group*.

We will refer to this definition as the classical definition.

EXAMPLE. In $\mathbb{R}^n$ consider the subgroup

$$\Gamma = \{k_1\alpha_1 + \cdots + k_n\alpha_n \colon k_1, \ldots, k_n \in \mathbb{Z}\},$$

where $\alpha_1, \ldots, \alpha_n$ is a basis in $\mathbb{R}^n$. This is a lattice isomorphic to $\mathbb{Z}^n$, and for this reason, it is called an *Abelian lattice* in $\mathbb{R}^n$. The group $\Gamma$ acts on this lattice by translations $x \to x + g$, $g \in \Gamma$. Therefore, the group $\mathbb{Z}^n = \Gamma$ is a crystallographic group for any $n = 1, 2, \ldots$ .

The above definitions are very general and work, for example, in the case of Lobachevsky space. We will consider the case where the ambient space is the Euclidean space $\mathbb{R}^n$. This is the only case that arises (so far) in the real-world physics in the description of crystals.

Crystallographic groups in Euclidean spaces may be defined as follows.

A subgroup $G$ of the group of motions of the $n$-dimensional Euclidean space is called an $n$-dimensional *crystallographic group* if its intersection with the group of translations $\mathbb{R}^n \subset \mathrm{E}(n)$ is an Abelian lattice $T \approx \mathbb{Z}^n$ in $\mathbb{R}^n$.

This definition is more convenient because its modification leads to a natural definition of quasi-crystallographic groups.

A subgroup $G$ of the group of motions of the space $\mathbb{R}^n$ is an $n$-dimensional *quasi-crystallographic group* in the sense of Novikov–Veselov if its intersection with the group of translations of $\mathbb{R}^n$ is a finitely generated subgroup $T$ that generates $\mathbb{R}^n$ as a vector space. When the group $T$ has the minimal possible rank equal to $n$, we obtain a crystallographic group.

We will show that in Euclidean spaces the two definitions of crystallographic groups are equivalent.

First we prove a general lemma on subgroups of the group $\mathrm{E}(n)$ of motions of the Euclidean space $\mathbb{R}^n$. Recall that every motion $g \in \mathrm{E}(n)$ is specified in Euclidean coordinates as

$$g = (A, b)\colon x \to Ax + b, \quad A \in \mathrm{O}(n),\ b \in \mathbb{R}^n.$$

**Lemma 6.12.** *Let $G$ be a subgroup of the group $\mathrm{E}(n)$.*

1. *The mapping*

$$\rho\colon G \to \mathrm{O}(n), \quad \textit{where} \quad g = (A, b) \to \rho(g) = A,$$

*is a homomorphism of groups. Its kernel is the normal subgroup $T \subset G$ consisting of all translations that belong to $G$.*

2. *If $A \in \rho(G)$ and $c \in T$, then $Ac \in T$.*

**Proof.** The first statement is obvious. We prove the second. If $A \in \rho(G)$, then there exists a transformation $g\colon x \to Ax + b$ that belongs to $G$. The inverse transformation $g^{-1}\colon x \to A^{-1}x - A^{-1}b$ also belongs to $G$. Consider the composition $g\tau g^{-1}$, where $\tau\colon x \to x + c$ is the translation by vector $c$:

$$x \to A(A^{-1}x - A^{-1}b + c) + b = x + Ac.$$

It is the translation by vector $Ac$, and it belongs to the group $T$. □

**Corollary 6.7.** *Let $G$ be a subgroup of the group $\mathrm{E}(n)$ and suppose that its intersection $T$ with the group of translations is an Abelian lattice. Then the quotient group $G/T$ is finite.*

**Proof.** Let $\alpha_1, \ldots, \alpha_n$ be the generators of the lattice $T$. They lie in a ball $D$ of finite radius with center $0 \in \mathbb{R}^n$, which contains only finitely many points of $T$. Each transformation $A \in \rho(G) \subset \mathrm{O}(n)$ is uniquely defined by the images $A(\alpha_1), \ldots, A(\alpha_n)$, which also belong to $D \cap T$. Hence if the group $G/T = \rho(G)$ were infinite, the set $D \cap T$ would consist of infinitely many points. But this contradicts the fact that $T$ is a lattice. Thus we conclude that the group $G/T$ is finite. □

If $G$ is a crystallographic group corresponding to a crystal $\Gamma$, then the image of the homomorphism $\rho$ is called the *group of symmetries of the crystal* or the *point group of the crystal* and is denoted by $S(\Gamma)$. Two point groups of $n$-dimensional crystals are equivalent if they are conjugate in $\mathrm{O}(n)$, i.e., there exists a transformation $g \in \mathrm{O}(n)$ such that $gS_1g^{-1} = S_2$. The kernel of the homomorphism $\rho$ is called the *crystal translation group* and is denoted by $T(\Gamma)$. For simplicity we will identify the elements of this group with translation vectors.

For Euclidean spaces the following theorem holds.

**Theorem 6.26.** *Let $G$ be a crystallographic group (according to the classical definition). Then its subgroup of translations $T$ is an Abelian lattice in $\mathbb{R}^n$ and has a finite index in the group $G$.*

**Proof.** We restrict ourselves to the case $n = 2$. By the following lemma, it suffices to prove the theorem for the case where the entire group $G$ consists of proper motions.

**Lemma 6.13.** *If $G$ is a crystallographic group in $\mathbb{R}^n$, then its normal subgroup $G_0$ formed by proper (preserving orientation) motions is crystallographic as well.*

**Proof.** If $G$ consists entirely of proper motions, then the lemma is obviously true. Let $g'$ be an improper motion in $G$. Then any improper motion $g \in G$ has the form $g = hg'$, where $h \in G_0$. The group $G_0$ preserves the lattice $\Gamma$, which can be represented as

$$\Gamma = \left(\bigcup_{i=1}^{k} G_0 x_i\right) \cup \left(\bigcup_{i=1}^{k} G_0(g'x_i)\right).$$

Therefore $G_0$ is also a crystallographic group. □

Recall that any proper motion of the plane is either a translation $x \to x + \alpha$ or a rotation about some point (Theorem 1.7).

A crystallographic group cannot contain rotations through arbitrarily small angles. Indeed, let $G$ contain a rotation through an angle $\varphi$ with center $x_0$. There exists a point of the crystal lying at a distance $\rho \leq R$ from $x_0$. Under rotation it turns into a point $x''$ lying at a distance $2\rho\sin(\varphi/2)$ from $x'$. By assumption, the distance between different points of the lattice is at least $r$, which implies the inequality $\sin\left(\frac{\varphi}{2}\right) \geq \frac{r}{2R}$ for the rotation angle $\varphi$ (see Figure 6.1).

Suppose that the group $G$ contains no translations. Then this group is isomorphic to the point group $S(\Gamma)$, which cannot contain arbitrarily small rotations and hence is finite. But since the lattice $\Gamma$ is infinite, the group $G$

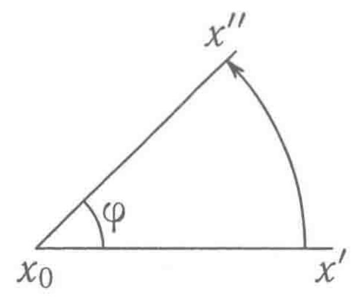


**Figure 6.1.** The rotation angle.

is also infinite, and we arrive at a contradiction. Therefore, the group $G$ contains at least one translation $x \to x + \beta_1$.

If the group $G$ does not contain rotations through angles other than multiples of $\pi$, then it necessarily contains a translation by a vector $\beta_2$ linearly independent of $\beta_1$. Otherwise the orbit of every point of the crystal would lie on a straight line with direction vector $\beta_1$, and so the crystal could not be a union of finitely many orbits. Now, let the group $G$ contain a rotation $A$ with center at a point $x_0$ through an angle $\varphi$ that is not a multiple of $\pi$. The vector $\beta_2 = A\beta_1$ is not proportional to $\beta_1$ and by Lemma 6.12 lies in $T(\Gamma)$.

Thus we have shown that the group of translations $T(\Gamma)$ contains translations by linearly independent vectors $\beta_1$ and $\beta_2$.

Now we will show that there exist two linearly independent vectors $\alpha_1$ and $\alpha_2$ whose linear combinations with integer coefficients form the translation lattice,

$$T(\Gamma) = \{k_1\alpha_1 + k_2\alpha_2,\ k_1, k_2 \in \mathbb{Z}\}. \tag{6.39}$$

First we choose in $T(\Gamma)$ a nonzero vector $\alpha_1$ of minimal length. Since $T(\Gamma)$ is imbedded in a lattice, such a vector does exist. Draw the straight line $\xi = t\alpha_1, t \in \mathbb{R}$, in the direction of $\alpha_1$ and find a translation vector $\alpha_2$, not proportional to $\alpha_1$, such that the point $\alpha_2 \in \mathbb{R}^2$ lies at the shortest distance to this line. Such a vector also exists. To prove this, introduce a basis $e_1$, $e_2$ in $\mathbb{R}^2$ such that $e_1 = \alpha_1$ and $e_2$ is orthogonal to $e_1$. Any vector of the lattice is representable as $x_1e_1 + x_2e_2$. We look for a vector in $T(\Gamma)$ with the least value of $|x_2| > 0$. Any vector of $T(\Gamma)$ can be translated by a vector $k\alpha_1$ with $k \in \mathbb{Z}$ so that it will have $0 \le x_1 \le 1$ with $x_2$ unchanged. Hence we search for a translation vector with minimum value of $|x_2|$ in a bounded domain containing finitely many points of the lattice. Therefore, such a vector $\alpha_2$ does exist.

Now we will show that the vectors $\alpha_1$ and $\alpha_2$ generate the lattice of translations. Let $\tau \in T(\Gamma)$. Since the vectors $\alpha_1$ and $\alpha_2$ form a basis in $\mathbb{R}^2$,

we have $\tau = x_1\alpha_1 + x_2\alpha_2$. Consider the fractional parts $\{x_1\}$ and $\{x_2\}$ of the numbers $x_1$ and $x_2$, so that $0 \le \{x_i\} < 1$, $i = 1, 2$. The vector $\tau' = \{x_1\}\alpha_1 + \{x_2\}\alpha_2$ also belongs to $T(\Gamma)$. If $\{x_2\} > 0$, then $\tau'$ does not lie on the line $t\alpha_1$, but is closer to it than $\alpha_2$, which contradicts the choice of $\alpha_2$. Therefore, $\{x_2\} = 0$. Now if $\{x_1\} > 0$, then the vector $\{x_1\}\alpha_1$ belongs to $T(\Gamma)$, but is of smaller length than $\alpha_1$. This contradicts the choice of $\alpha_1$. Therefore, $x_1, x_2 \in \mathbb{Z}$. Thus we have proved equality (6.39).

It remains to prove that the point group of the crystal $S(\Gamma) = G/T(\Gamma)$ is finite. This follows from Corollary 6.7.

The proof of Theorem 6.26 for two-dimensional crystals is completed. □

This theorem implies that in case of Euclidean spaces the two definitions of crystallographic groups given above are equivalent.

Although a crystal consists of points, it is expedient to associate with it a partition of the space into identical parallelepipeds. Namely, if $\alpha_1, \dots, \alpha_n$ are the generators of the group of translations of a crystal in $\mathbb{R}^n$, then the parallelepiped spanned by these vectors is called a *primitive cell* of the crystal (see Figure 6.2). Each cell contains finitely many points of the crystal. Take a vertex of some cell for the origin and denote by $\xi_1, \dots, \xi_k$ the radius-vectors of the points of the lattice lying inside the primitive cell, i.e., satisfying the inequalities

$$\xi_i = a_{i1}\alpha_1 + \cdots + a_{in}\alpha_n, \quad 0 \le a_{ij} < 1,\ i = 1, \dots, k,\ j = 1, \dots, n.$$

Obviously, the translation lattice and the vectors $\xi_1, \dots, \xi_k$ completely specify the lattice.

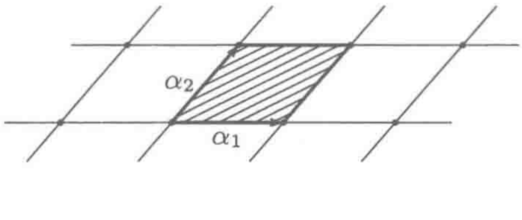


**Figure 6.2.** A primitive cell.

In the simplest case, where $k = 1$ and $\xi_1 = 0$, all the points of the crystal lie at the points of the translation lattice $k_1\alpha_1 + \cdots + k_n\alpha_n$. Such crystals are called the *Bravais lattices*.

From the affine point of view, all the Bravais lattices in $\mathbb{R}^n$ are equivalent to each other because any two of them can be superposed by an affine transformation. However, they may have different symmetry groups, which is connected with arithmetic properties of their generators.

There is another partition of the space into cells (of possibly different shape), which is used in physics. Namely, to each point of the lattice (atom)

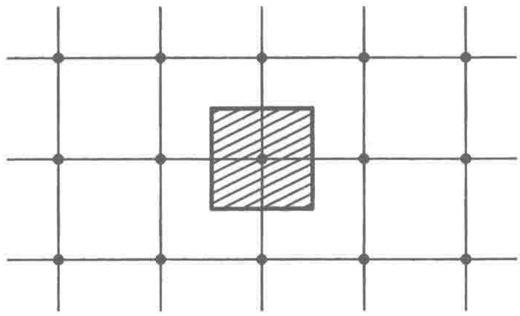

**Figure 6.3.** A Wigner–Seitz cell.

one assigns the set of points which lie closer to this atom than to any other one (see Figure 6.3). These cells are called the *Wigner–Seitz cells* (in number theory, they are called the *Dirichlet domains*).

Now we proceed to the classification of crystallographic groups in the two-dimensional space. First we prove the following general theorem.

**Theorem 6.27.** *The point groups of crystals in $\mathbb{R}^n$ are exactly the groups of (orthogonal) symmetries of Abelian lattices $T$ in $\mathbb{R}^n$.*

**Proof.** By Lemma 6.12, if $A \in S(\Gamma)$ and $c \in T(\Gamma)$, then $Ac \in T(\Gamma)$. Hence each element of the point group of the crystal determines a symmetry of the translation lattice $T(\Gamma)$. Conversely, if $S' \subset \mathrm{O}(n)$ is the group of symmetries of some $n$-dimensional lattice $T'$ in the translation group of $\mathbb{R}^n$, then the group $G$ generated by translations by the vectors in $T'$ and by elements of $S'$, is a crystallographic group. One may take the lattice $T'$ for the crystal; then the point group of the crystal will be the group $S'$. □

The point groups of $n$-dimensional crystals are called *crystallographic classes*. It is assumed that the conjugate groups in $\mathrm{O}(n)$ (i.e., equivalent groups) determine the same class. By Theorem 6.26, they are finite subgroups of $\mathrm{O}(n)$, but not every such subgroup can be realized as a point group of some lattice.

The complete classification of two-dimensional crystallographic classes is given by the following theorem.

**Theorem 6.28.** *If $S \subset \mathrm{O}(2)$ is the point group of a crystal in $\mathbb{R}^2$, then it is one of the following groups:*

1. *The cyclic group $C_k = \mathbb{Z}_k$, where $k = 1, 2, 3, 4, 6$, generated by rotations through the angle $2\pi/k$.*

2. *The dihedral group $D_k$, where $k = 1, 2, 3, 4, 6$, generated by rotations in $C_k$ and reflections in a straight line.*

*In case 1 the crystallographic group $G$ contains only proper motions; in case 2 it contains improper motions as well.*

Recall that the *dihedral group* $D_k$ is the group of motions of the plane which take the regular $k$-gon into itself, i.e., the group of symmetries of this $k$-gon. The group $D_k \subset \mathrm{O}(2)$ is generated by rotation through the angle $\frac{2\pi}{k}$ about the center of symmetry and by reflection in a symmetry axis of the $k$-gon. Therefore, $D_k$ contains, as a normal subgroup of index 2, the cyclic group $C_k$ generated by rotation through the angle $\frac{2\pi}{k}$. The subgroup $C_k$ is the kernel of the homomorphism $\det\colon D_k \to \{\pm 1\}$.

**Proof of Theorem 6.28.** Assume that the group $G$ consists only of proper transformations. By Theorem 6.26, the point group $S(\Gamma)$ of the crystal is a finite subgroup of the group $\mathrm{SO}(2)$ of rotations of the plane. The group $\mathrm{SO}(2)$ is isomorphic to the group of complex numbers of modulus one, $\mathrm{SO}(2) \simeq \{e^{i\varphi}\}$, where rotation through an angle $\varphi$ corresponds to multiplication by $e^{i\varphi}$. This group is commutative and its finite subgroups are cyclic and generated by rotations through the angle $\frac{2\pi}{k}$, $k = 1, 2, \dots$.

In the translation group $T(\Gamma)$, select a basis $\alpha_1$, $\alpha_2$ such that $\alpha_1$ has the minimum length among nonzero vectors of $T(\Gamma)$.

Let $g_k$ denote the rotation through the angle $\frac{2\pi}{k}$, which is the generator of the group $C_k = \mathbb{Z}_k$. If $k > 6$, then it can easily be verified that $|\alpha_1 - g_k(\alpha_1)| < |\alpha_1|$, which contradicts the choice of $\alpha_1$. Therefore, if $S(\Gamma) = C_k$, then $1 \le k \le 6$.

A rotation through the angle $\pi$ is always a symmetry of the of translation group $T(\Gamma)$. If the group $C_5 = \mathbb{Z}_5$ is the point group of the lattice, then, combined with rotation through the angle $\pi$, it generates the group $C_{10}$ of rotations through the angles $\frac{\pi k}{5}$, specifying symmetries of the lattice $T(\Gamma)$. But this contradicts the fact that $C_k$ may be imbedded into the point group of a crystal only for $k \le 6$. Therefore, the group $C_5$ cannot be realized as the point group of a two-dimensional crystal.

We leave it as an exercise for the reader to show that the remaining groups $C_k$, where $k = 1, 2, 3, 4, 6$, can be realized as point groups of crystals in $\mathbb{R}^2$.

It remains to consider the case where the group $G$ contains improper motions. In this case the point group $S$ contains the subgroup $S_0$ of index two formed by proper symmetries, which, according to the above, is $C_k$ with $k$ equal to 1, 2, 3, 4, or 6.

Any improper orthogonal transformation of the plane is the reflection $\sigma$ in some straight line (this can easily be proved or deduced from Theorem 1.7). It is easily seen that $\sigma$ and the elements of $S_0 = C_k$ generate

the dihedral group $D_k$. We again leave it as an exercise to prove that any dihedral group $D_k$ for $k = 1, 2, 3, 4, 6$ can be realized as a point group of a crystal in $\mathbb{R}^2$. $\square$

This theorem does not say how dihedral groups are realized as groups of automorphisms (linear invertible mappings) of translation lattices. The automorphisms of the Abelian lattice $T \subset \mathbb{R}^n$ are in one-to-one correspondence with matrices of the group $\mathrm{GL}(n, \mathbb{Z})$, i.e., invertible $n \times n$ matrices with integer coefficients. Indeed, if $\alpha_1, \dots, \alpha_n$ is a basis of the lattice $T$, which is mapped by the automorphism $f \colon T \to T$ into $f(\alpha_i) = a_i^j \alpha_j$, $i = 1, \dots, n$, then the matrix $(a_i^j)$ lies in $\mathrm{GL}(n, \mathbb{Z})$ and uniquely determines the automorphism $f$. Each element $g \in S$ of the point group $S$ specifies an automorphism $T \to T$, and we obtain a homomorphic imbedding

$$S \to \mathrm{GL}(n, \mathbb{Z}).$$

Two point groups (groups of symmetries) of a lattice are said to be equivalent if they are conjugate as subgroups of $\mathrm{GL}(n, \mathbb{Z})$. It turns out that the groups $D_1$, $D_2$, and $D_3$ admit two nonequivalent imbeddings each. Hence we have the following result.

**Theorem 6.29.** *There are* 13 *pairwise nonequivalent point groups of two-dimensional lattices, five of which consist of proper transformations only, while the remaining eight contain also improper transformations.*

Now, in order to obtain the classification of crystallographic groups in $\mathbb{R}^2$, we must find all groups $G$ such that $G$ contains the normal subgroup $T = \mathbb{Z}^2$ and the quotient $S = G/T$ is a two-dimensional crystallographic class. Each such group is imbedded in a natural way into the group $\mathrm{E}(2)$ of motions of the plane: any element of $G$ is representable as $g = t \cdot s$, where $t \in T$ and $s \in S \subset \mathrm{O}(2)$, and it corresponds to the motion $t \cdot s$ in $\mathrm{E}(2)$. Note that the elements of $S$ themselves need not belong to $G$ arising only in combinations $g = t \cdot s$ with nontrivial translations $t \neq 0$.

Two crystallographic groups $G_1$ and $G_2$ are *equivalent* if there exists a proper affine transformation $h \colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $hG_1h^{-1} = G_2$.

For two-dimensional crystallographic groups, equivalence coincides with the existence of a usual algebraic isomorphism $G_1 \to G_2$.

We state the following theorem for the sake of completeness, omitting the proof.

**Theorem 6.30.** *There are* 17 *pairwise nonequivalent two-dimensional crystallographic groups, five of which contain only proper transformations, and* 12 *contain also improper transformations.*

Thus we conclude that there are three levels of classification of planar crystals, which we list in the order of increasing accuracy:

1) By equivalence classes (conjugacy in $\mathrm{O}(n)$) of point groups, i.e., by crystallographic classes, — 10 types.

2) By conjugacy classes of point groups in $\mathrm{GL}(2,\mathbb{Z})$ — 13 types.

3) By equivalence classes (conjugacy in $\mathrm{E}(n)$) of crystallographic groups — 17 types.

Two-dimensional crystallographic groups are equivalent if and only if they are isomorphic as groups. For three-dimensional groups this is not the case.

In dimensions 2 and 3 the classification of all crystallographic groups was obtained independently by Fedorov and Schoenflies; hence two- and three-dimensional crystallographic groups are also called *Fedorov's groups*.

Now we proceed to classification of three-dimensional crystallographic groups.

Recall that any proper orthogonal transformation $A$ of $\mathbb{R}^3$ is a rotation about some axis (Lemma 1.5).

All cyclic groups $C_k = \mathbb{Z}_k$ are realized as groups of rotations through the angles $2\pi l/k$, $l = 1, \dots, k$, about some axis in $\mathbb{R}^3$.

The action of the dihedral group $D_k$ in the coordinate plane $x^3 = 0$ can be extended to a proper action in $\mathbb{R}^3$: the cyclic subgroup $C_k$ acts by rotations about the axis $Ox^3$, and the reflection in the line $x^2 = 0$ in the plane $x^3 = 0$ is extended to rotation of $\mathbb{R}^3$ through the angle $\pi$ about this line.

Note that the group $D_1$ acts as the group $C_2$.

Besides the groups $C_k$ and $D_k$ we must introduce the groups of proper symmetries of regular polyhedra. Namely, a polyhedron is said to be regular if for each pair of tuples $(P_1, Q_1)$ and $(P_2, Q_2)$, where $P_i$ is a vertex of the polyhedron and $Q_i$ is an edge with one of the end-points at $P_i$, there exists a proper rotation, which is a symmetry of the polyhedron and takes $(P_1, Q_1)$ into $(P_2, Q_2)$. There are five regular polyhedra: tetrahedron, cube, octahedron, dodecahedron, and icosahedron (see Figure 6.4). Denote their proper groups of symmetries by $T$ (for tetrahedron), $O$ (for cube and octahedron; these groups are the same), and $Y$ (for dodecahedron and icosahedron; their groups of symmetries also coincide).

The tetrahedral group $T$ is isomorphic to the group $A_4$ formed by even permutations of four elements. The octahedral group $O$ is isomorphic to the symmetric group $S_4$ formed by all permutations of four elements. Hence it contains $T$ as a normal subgroup of index 2, which is the kernel of the

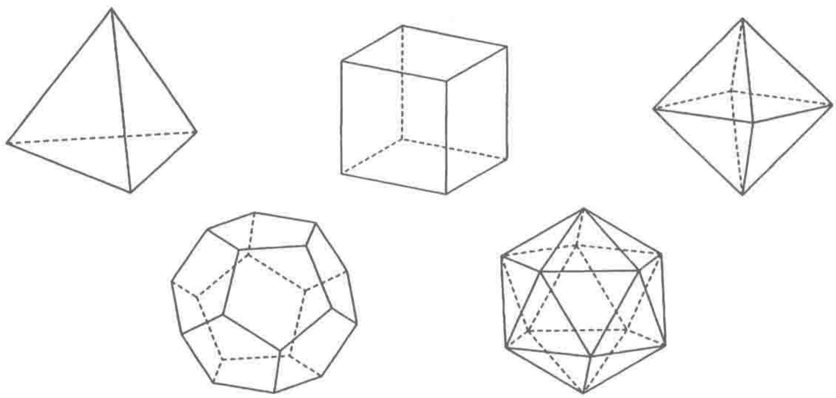

**Figure 6.4.** Tetrahedron, cube, octahedron, dodecahedron, icosahedron.

homomorphism sgn: $S_4 \to \{\pm 1\}$. The icosahedral group $Y$ is isomorphic to the group $A_5$ of even permutations of five elements. Therefore, their orders are

$$|T| = \frac{4!}{2} = 12, \qquad |O| = 4! = 24, \qquad |Y| = \frac{5!}{2} = 60.$$

We have the following result.

**Theorem 6.31.** *A finite group $G$ consisting of proper rotations of $\mathbb{R}^3$ is isomorphic to one of the groups in the following list:*

1) *Cyclic groups $C_k$, $k = 1, 2, \ldots$.*
2) *Dihedral groups $D_k$, $k = 2, 3, \ldots$.*
3) *Tetrahedral, octahedral, and icosahedral groups $T$, $O$, and $Y$.*

**Proof.** We will outline the proof, presenting only the main ideas. Let $G$ be a finite subgroup of SO(3) and let $N = |G|$ be its order (the number of its elements).

Any proper rotation of space is a rotation about some axis (Lemma 1.5). Let $S^2$ be the sphere of unit radius with center at the origin in $\mathbb{R}^3$. It is mapped into itself by transformations in $G$.

We will say that a point $p \in S^2$ is a pole if there exists a nontrivial rotation $A \in G$ keeping $p$ fixed. A pole $p$ is of multiplicity $\nu_p$ if the group $G$ contains exactly $\nu_p$ different transformations with pole at $p$. This means that there is a cyclic subgroup $H = Z_{\nu_p} \subset G$ consisting of rotations through the angles $2\pi k/\nu_p$, where $k = 1, \ldots, \nu_p$, about the axis passing through the points 0 and $p$.

Now the group $G$ falls into the union of left cosets $gH$ by the subgroup $H$:

$$G = \bigcup_{i=1}^{k} (g_i H),$$

where $g_iH$ is the set of elements of $G$ representable as $g_ih$, $h \in H$. The cosets are disjoint, $g_iH \cap g_jH = \emptyset$ for $i \neq j$. Hence $N = k|H| = k\nu_p$. Therefore, the multiplicity of each pole is a divisor of the order $N$ of the group $G$. The number $k_p = N/\nu_p$ is equal to the number of points in the orbit of $p$. Each point $gp$ of the orbit is a pole of the transformation $gAg^{-1} \in G$. It is easily seen that this pole is of the same multiplicity as the pole $p$, since the point $gp$ stays fixed under transformations in the group $gAg^{-1}$.

Let us count the number of pairs $(A, p)$, where $A$ is a nontrivial rotation in $G$ and $p$ is a pole of $A$. Since each rotation has two poles, the number of such pairs is $2(N-1)$. On the other hand, the contribution of each pole into this sum is $(\nu_p - 1)$. Thus we obtain

$$2(N-1) = \sum k_p(\nu_p - 1), \tag{6.40}$$

where the sum in the right-hand side is extended over different orbits of poles (rather than over all poles). Since $N = k_p\nu_p$ for each pole $p$, dividing both sides of this relation by $N$ we obtain

$$2 - \frac{2}{N} = \sum\Big(1 - \frac{1}{\nu_p}\Big).$$

The subsequent proof of the theorem consists in analyzing this relation. If $N > 1$, all the solutions to this relation are exhausted by the following list:

1) $\nu_1 = \nu_2 = N$ and $N$ is arbitrary (two orbits);
2) $\nu_1 = \nu_2 = 2, \nu_3 = n$, $N = 2n$ and $n$ is arbitrary (three orbits);
3) $\nu_1 = 2, \nu_2 = 3, \nu_3 = 3, N = 12$ (three orbits);
4) $\nu_1 = 2, \nu_2 = 3, \nu_3 = 4, N = 24$ (three orbits);
5) $\nu_1 = 2, \nu_2 = 3, \nu_3 = 5, N = 60$ (three orbits).

All these cases can be realized and correspond to the following groups: 1) cyclic groups $C_N$; 2) dihedral groups $D_n$; 3) tetrahedral group $T$; 4) octahedral group $O$; 5) icosahedral group $Y$.

We omit the cumbersome arguments leading to this conclusion and stop here the proof of the theorem. □

Reflection in three-dimensional space is specified by the matrix $-1$; it changes orientation and commutes with any rotation. Therefore, the group $\mathrm{O}(3)$ is the direct product:

$$\mathrm{O}(3) = \mathrm{SO}(3) \times \{\pm 1\}.$$

Any group $G$ of proper rotations may be extended to the group $G \times \{\pm 1\}$ by adding the elements $-g$, where $g \in G$. There is another method of

constructing subgroups of O(3) from subgroups of SO(3): let a group $G \subset$ SO(3) contain a subgroup $G_0$ of index two. Denote by $GG_0$ the group of elements $g$ and $-g'$, where $g \in G_0$ and $g' \in G \setminus G_0$.

The classification of three-dimensional crystallographic classes is given by the following theorem.

**Theorem 6.32.** 1. *It the point group of a three-dimensional crystal contains only proper rotations, then it belongs to the following list:*

*cyclic groups* $C_k$, $k = 1, 2, 3, 4, 6$;

*dihedral groups* $D_k$, $k = 2, 3, 4, 6$;

*tetrahedral and octahedral groups* $T$ *and* $O$.

2. *If the point group of a three-dimensional crystal contains improper motions, then it belongs to the following list:*

$C_k \times \{\pm 1\}$, *where* $k = 1, 2, 3, 4, 6$;

$D_k \times \{\pm 1\}$, *where* $k = 2, 3, 4, 6$;

$C_2C_1, C_4C_2, C_6C_3, D_4D_2, D_6D_3, D_2C_2, D_3C_3, D_4C_4, D_6C_6$;

$OT$, $T \times \{\pm 1\}$, $O \times \{\pm 1\}$.

*All the groups of these lists are realized as three-dimensional crystallographic classes.*

Therefore, there are 32 different 3-dimensional crystallographic classes. The following theorem shows how complicated is a finer classification of 3-dimensional crystals.

**Theorem 6.33.** *There are* 72 *pairwise nonequivalent point groups of* 3-*dimensional lattices and* 230 *different* 3-*dimensional crystallographic groups,* 65 *of which contain only proper motions and* 165 *also contain improper motions.*

We may weaken the notion of equivalence and regard two crystallographic groups as equivalent if they are conjugate to an affine (not necessarily proper) transformation. In these terms there are 219 different three-dimensional crystallographic groups. In the two-dimensional case this difference in definitions has no influence, and the crystallographic groups are equivalent if they are simply isomorphic.

**6.2.2. Quasi-crystallographic groups.** At present, the science and technology deal with many substances taken from nature or produced artificially for technical purposes. Some of them closely resemble a solid body, i.e., they fit well the definition of a lattice (see above), which, however, is not regular. This means that their properties are not periodic with (free Abelian)

group $\mathbb{Z}^n$, which generates an Abelian lattice in Euclidean space $\mathbb{R}^n$ (in reality, modern physics deals with the cases $n = 1, 2, 3$).

At the same time, according to the presently adopted conceptions, some of these substances, discovered in the 1980s, seem to have a continuous long-range correlation between positions of atoms and a symmetry which is impossible for crystals. The properties of the most important substances of this kind are similar to those of crystalline solid bodies; they are described by so-called "quasi-periodic functions" in $\mathbb{R}^n$.

A function $f(x^1, \dots, x^n)$ is called *quasi-periodic* relative to the dual quasi-lattice generated by vectors $\omega_1, \dots, \omega_{n+k} \in \mathbb{R}^{n*}$ if it can be expanded in a trigonometric series

$$f(x^1, \dots, x^n) = \sum_{m \in \mathbb{Z}^{n+k}} a_m \exp\left( i \sum_{j=1}^{n+k} m_j \langle \omega_j, x \rangle \right),$$

where

$$x = (x^1, \dots, x^n), \qquad m = (m_1, \dots, m_{n+k}),$$

$$\langle \omega_j, x \rangle = \sum_{q=1}^{n} \omega_{jq} x^q, \qquad \omega_j = (\omega_{j,1}, \dots, \omega_{jn}), \qquad j = 1, \dots, n+k.$$

The lattice of points $y_\alpha \in \mathbb{R}^n$ is called *quasi-periodic* if the function

$$f_\varphi(x) = \sum_\alpha \varphi(x - y_\alpha)$$

is quasi-periodic for each smooth function $\varphi(x)$ with compact support in $\mathbb{R}^n$.

A quasi-periodic function can always be represented as a function of $N$ variables, $N = n + k$, that is periodic relative to a regular lattice in $\mathbb{R}^N$ with periods $T_p \in \mathbb{R}^N$, $p = 1, \dots, N$, restricted to an irrational $n$-dimensional direction, $f(x^1, \dots, x^n) = \Phi(y^1, \dots, y^N)$, where

$$y^j = y_0^j + a_q^j x^q, \quad y = (y^1, \dots, y^N),$$
$$\Phi(y + T_p) = \Phi(y).$$

Quasi-periodic lattices suitable to our purposes are constructed as follows: suppose we are given a regular crystallographic lattice of points in $\mathbb{R}^N$, a subspace $\mathbb{R}^n \subset \mathbb{R}^N$, a vector $\alpha \in \mathbb{R}^n$, and a bounded convex body $Y$ in $\mathbb{R}^N$. Consider the tubular domain

$$S_\alpha = \{\alpha + \mathbb{R}^n + Y\} \subset \mathbb{R}^N.$$

We will project all the points of the lattice lying inside $S_\alpha$ to the subspace $\mathbb{R}^n$ along some direction $\mathbb{R}^\perp$, where

$$\mathbb{R}^\perp \oplus \mathbb{R}^n = \mathbb{R}^N.$$

Usually the lattice is taken to consist of all integer-valued vectors of $\mathbb{R}^N$ in some basis $(e_1, \ldots, e_N)$, while $\mathbb{R}^n$ is taken to be an irrational subspace. In this way we obtain a set of points in $\mathbb{R}^n$ which form a quasi-lattice. It is pointed out in the survey [**58**] that the equivalence of this definition to the initial definition was carefully verified by V. I. Arnold. We will not prove it here.

However, the simplest and most interesting quasi-periodic point lattices are constructed by means of so-called *tilings* of the space $\mathbb{R}^n$ with finitely many polyhedra $K_1, \ldots, K_L$, which adjoin each other in the right way (i.e., by a complete face of some dimension) and cover the entire space $\mathbb{R}^n$ so that their interiors are disjoint. The famous Penrose tiling of $\mathbb{R}^2$ is specified by two rhombs $K_1$, $K_2$ with acute angles 72° and 36° and sides of equal length. We will describe it below. As pointed out by de Bruijn in 1981, this tiling is quasi-periodic, i.e., the set of vertices forms a quasi-periodic lattice. This pair of rhombs enable one to construct a large family of Penrose-type tilings following certain *local rules* (see below), by which only quasi-periodic tilings are obtained.

Taking into account the discovery in the 1980s of substances which are apparently the *quasi-crystals* in this sense, the idea of local rules that impose quasi-periodicity, may be regarded as an explanation (to some extent) of this form of long-range regularity in atom positioning. This subject was elaborated in depth by a number of physicists (L. Levitov, A. Katz, and others), and taken also by geometers (see [**58**]).

We present here only the simplest example of a tiling generated by the two Penrose rhombs mentioned above. We introduce the "colored rhombs" according to Figures 6.5 and 6.6. We will require the vertices in a tiling to be of one of the six types shown in Figure 6.7. (These are our local rules.)

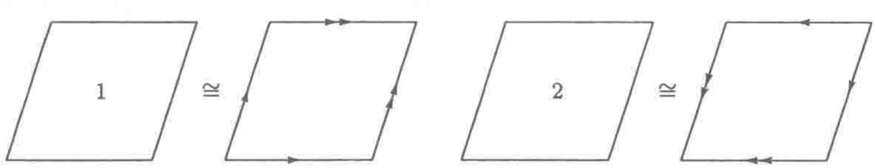


**Figure 6.5.** Colored rhombs $K_1$.

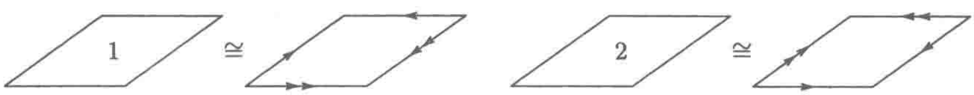


**Figure 6.6.** Colored rhombs $K_2$.

The tilings will be composed of colored rhombs (see above) and of the same rhombs turned through multiples of $72° = 2\pi/5$. In this way the

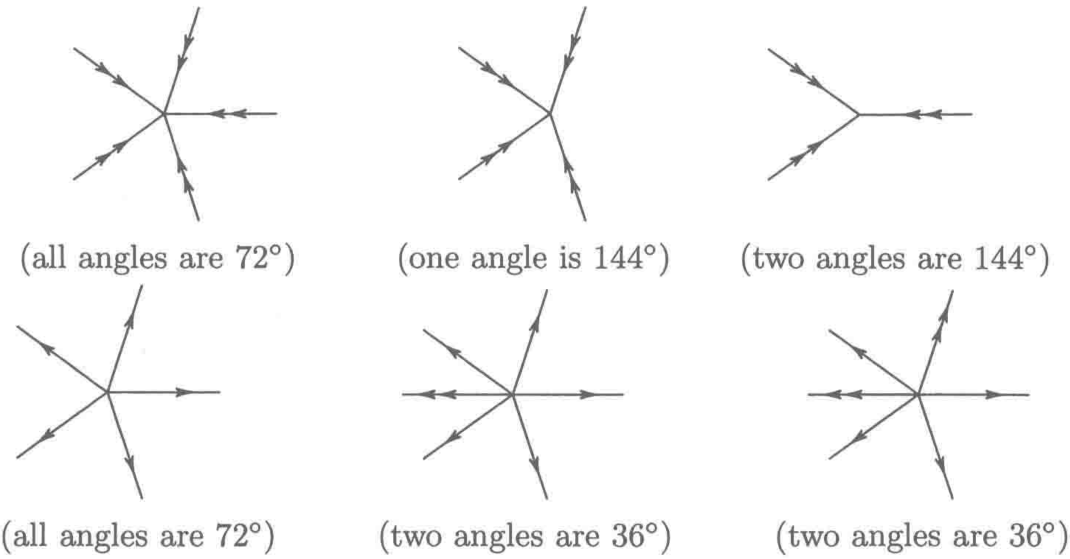


**Figure 6.7.** Local rules.

initial four colored rhombs give rise to 20 figures. We will allow only parallel translations of these 20 colored rhombs. We state without proof the following important fact.

**Theorem 6.34.** *Any tiling of the plane by colored rhombs satisfying the local rules (the choice of six admissible vertices) is quasi-periodic.*

When gluing the rhombs to each other, they must adjoin along sides of the same color. Let us emphasize once more that the positions in $\mathbb{R}^2$ of the colored rhombs $(K_1)$, $(K_2)$ are shown in Figures 6.5 and 6.6 up to parallel translations and rotations through multiples of 72°.

A certain lack of uniqueness, the possibility to sometimes make a choice in combining the rhombs, is the reason that we have a continuum of different tilings. However, this arbitrariness in gluing is so limited that it restricts us to only quasi-periodic tilings. This reflects the long-range order in atom positioning.

Now we turn to the group-theoretic aspect of quasi-crystallographic tilings dealing with their symmetries. The above example shows that, besides the group of translations, there is a cyclic group of order 5 of rotations of the plane, built in the structure of all these tilings, which did not appear in crystallographic groups of the plane. Let us construct these tilings following de Bruijn's scheme. Consider the Euclidean space $\mathbb{R}^5$ with orthonormal basis $(e_1, e_2, e_3, e_4, e_5)$ and the action of the group of order 5 by the formula $T(e_j) = e_{j+1}$, where $j = 1, 2, 3, 4, 5$ and the number $j$ is treated as the residue modulo 5. The space $\mathbb{R}^5$ decomposes into the direct sum of

orthogonal spaces

$$\mathbb{R}^5 = \mathbb{R}^2_1 \oplus \mathbb{R}^2_2 \oplus \mathbb{R}^1,$$

where $\mathbb{R}^1$ is generated by the vector

$$e = e_1 + \cdots + e_5,$$

$\mathbb{R}^2_1$ is generated by the vectors $e_\pm$,

$$e_+ = \sum_{n=1}^{5} \cos\Bigl(\frac{2\pi n}{5}\Bigr) e_n, \qquad e_- = \sum_{n=1}^{5} \sin\Bigl(\frac{2\pi n}{5}\Bigr) e_n,$$

and $\mathbb{R}^2_2$ is generated by the vectors $a_\pm$,

$$a_+ = \sum_{n=1}^{5} \cos\Bigl(\frac{4\pi n}{5}\Bigr) e_n, \qquad a_- = \sum_{n=1}^{5} \sin\Bigl(\frac{4\pi n}{5}\Bigr) e_n.$$

It is easily seen that, over the field of complex numbers, the vectors $e_+ \pm ie_-$ and $a_+ \pm ia_-$ are the eigenvectors of the operator $T\colon \mathbb{R}^5 \to \mathbb{R}^5$ with eigenvalues $\exp(\pm 2\pi i/5)$ and $\exp(\pm 4\pi i/5)$, respectively. Therefore, the transformation $T$ acts in the planes $\mathbb{R}^2_1$ and $\mathbb{R}^2_2$ as rotation through the angles 72° and 144°, respectively. These subspaces are orthogonal to each other. Consider the 5-dimensional unit cube $I^5$ generated by the vectors $e_1, e_2, e_3, e_4, e_5$ in the space $\mathbb{R}^5$. Denote by $S_\alpha$ the body in $\mathbb{R}^5$ swept out by all shifts of this cube by the vectors $\gamma \in \mathbb{R}^2_1 + \alpha$:

$$S_\alpha = \bigcup_\gamma \{I^5 + \gamma\}.$$

Suppose that the vector $\alpha$ is *regular*, i.e., the boundary of $S_\alpha$ does not contain integer-valued vectors (points of the cubic lattice in $\mathbb{R}^5$).

The projections of two-dimensional faces of the cube $I^5$ on the plane $\mathbb{R}^2_1$ give the required Penrose's rhombs. The projections of all 2-faces of $I^5$ on $\mathbb{R}^2_1$ (along $\mathbb{R}^2_2 \oplus \mathbb{R}$) in their totality produce just the Penrose–de Bruijn tilings dependent on the vector $\alpha$. Quasi-periodicity of this construction is obvious.

This is the general "method of projections" for constructing tilings: we take the Euclidean space $\mathbb{R}^N$ with orthonormal basis $(e_1, \dots, e_N)$ and the cubic lattice generated by these vectors. Suppose we are given a subspace $\mathbb{R}^n \subset \mathbb{R}^N$ and a vector $\alpha \in \mathbb{R}^{N-n}$, where $\mathbb{R}^{N-n}$ is orthogonal to $\mathbb{R}^n$. Define the body $S_\alpha$ as the union

$$S_\alpha = \bigcup_\gamma (I^N + \gamma + \alpha),$$

where $\gamma$ runs over $\mathbb{R}^n$. The projections of the $n$-dimensional faces of the cube $I^N$ on the subspace $\mathbb{R}^n$ provide the elements that form the tiling.

If the vector $\alpha$ is regular, as above, then the set of all $n$-dimensional faces of integer-valued cubes $I^N$ in $S_\alpha$, projected to $\mathbb{R}^n$, provides the required tiling, which is quasi-periodic by construction.

In examples of this type it is natural to associate with this construction the subgroups of the multidimensional (cubic) crystallographic group in $\mathbb{R}^N$ related somehow to the subspace $\mathbb{R}^n \subset \mathbb{R}^N$, although these groups do not translate the points of the lattice into one another.

We follow the definition of quasi-crystallographic groups given in the previous section, which relied on geometry of the initial "physical" Euclidean space: a quasi-crystallographic group $G$ is an arbitrary subgroup of the group of isometries (motions) of the Euclidean space $\mathbb{R}^n$,

$$G \subset \mathrm{E}(\mathbb{R}^n),$$

for which the intersection with the group of translations, $G \cap \mathbb{R}^n = \mathbb{Z}^N$, is a finitely generated Abelian group generating the entire subspace $\mathbb{R}^n$ over $\mathbb{R}$. Thus $N \geq n$. We have shown in the previous section that for $N = n$ this yields an equivalent definition of ordinary crystallographic groups. For $N > n$ we arrive at a new type of groups describing the symmetry of crystals. These groups of symmetries cannot be seen directly from the positioning of atoms in $\mathbb{R}^n$; it is an extension $G$ of the group of frequency vectors $\omega_j$, $j = 1, \ldots, N$,

$$\mathbb{Z}^N = \left\{\sum m_j \omega_j\right\} \subset \mathbb{R}^{n*}, \quad m_j \in \mathbb{Z},$$

as a subgroup of dual translations $\omega_j \in \mathbb{R}^{n*}$ in the space of wave vectors. We require the equality $G \cap \mathbb{R}^{n*} = \mathbb{Z}^N$ to hold, i.e., we assume that the extension does not increase the amount of translations.

The simplest example is the group discussed above, where $n = 2$ and the basis frequencies are shown in Figure 6.8 with $\omega_{j+1} = e^{2\pi i/5}\omega_j$.

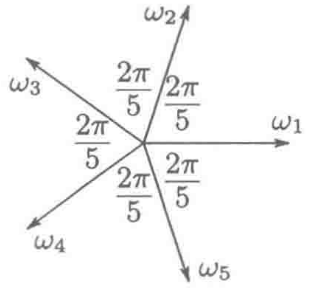


**Figure 6.8**

We have $\sum_{j=1}^{5} \omega_j = 0$. Hence the vectors $\omega_j$ generate the subgroup $\mathbb{Z}^4 \subset \mathbb{R}^{2*}$. The group $G$ has two generators $T$, $\omega_0$, where $T^5 = 1$, $T\omega_j T^{-1} = \omega_{j+1}$, and $\omega_0$, $\omega_1$, $\omega_2$, $\omega_3$ commute with each other. The quotient group

$G/(G\cap\mathbb{R}^n)$ by the subgroup of translations is a subgroup of the orthogonal group

$$G/(G\cap\mathbb{R}^n)\subset O_n(\mathbb{R}).$$

In the above example this group is isomorphic to the group $\mathbb{Z}/5\mathbb{Z}$, i.e., to the cyclic group of order 5. A more general class of examples obtained from the above construction of quasi-crystallographic groups in multidimensional Euclidean spaces $\mathbb{R}^N$, which preserve an irrational direction $\mathbb{R}^n\subset\mathbb{R}^N$, always lead to quasi-crystallographic groups $G$ such that the quotient group

$$G/(G\cap\mathbb{R}^{n*})\subset O_n(\mathbb{R})$$

is finite. Note that the Novikov–Veselov definition (see the previous section) does not necessarily imply this property: already for $n=2$ there are fairly simple examples where this quotient group in infinite, i.e., there appear rotations through irrational angles. The classification theory of such groups is developed in [**58**]. We will present here the simplest facts about them. Consider the case $n=2$.

**Theorem 6.35.** *An angle $\varphi$ determines a rotation in a two-dimensional quasi-crystallographic group if and only if the number $z_0=e^{i\varphi}$ is an algebraic integer, i.e., there is a polynomial $P(z)=z^N+\sum_{j=1}^N a_jz^{N-j}$, $a_N=1$, $a_j=a_{N-j}$, $a_j\in\mathbb{Z}$, $i=1,\dots,n$, such that $P(z_0)=0$.*

**Proof.** *Sufficiency.* Define the group $G$ with generators $t=1\in\mathbb{C}=\mathbb{R}^2$ and $z_0=e^{i\varphi}\in U_1$. Consider an irreducible polynomial $P(z)$ with integer coefficients such that $P(z_0)=0$. Let

$$P(z)=z^N+a_1z^{N-1}+\cdots+a_N,\quad a_j\in\mathbb{Z}.$$

Then all powers $z_0^M$, $M\geq N$, can be expressed in terms of the basis $\omega_1=1$, $\omega_2=z_0,\dots,\omega_N=z^{N-1}$. We know that $P(\bar z_0)=P(z_0^{-1})=0$. At the same time

$$z_0^NP\Big(\frac{1}{z_0}\Big)=P(z_0)=0.$$

Hence we conclude that $a_k=a_{N-k}$ and

$$z_0^{-1}=-(z_0^{N-1}+a_{N-1}z_0^{N-2}+\cdots+a_2z_0+a_1).$$

Therefore, the element $\bar z_0=z_0^{-1}$ is also expressible in terms of the same basis with integer coefficients. Thus we see that $G$ is a quasi-crystallographic group.

*Necessity.* Let $\varphi$ be a rotation in a quasi-crystallographic group $G$. Consider the subgroup generated by $z_0=e^{i\varphi}$ and translation $t\in\mathbb{C}$. The vectors $z_0^kt$, $k\in\mathbb{Z}$, are translations that belong to our subgroup. The existence of a finite $\mathbb{Z}$-basis in $G$ implies that there is a number $M$ such that the vectors $z^kt$ with $|k|<M$ generate the Abelian group generated by all the vectors $\{z^qt,q\in\mathbb{Z}\}$. Therefore, the number $z_0$ is a root of a polynomial

$P(z_0, z_0^{-1}) = 0$ with integer coefficients and the leading coefficient equal to one. Multiplying by a suitable power of the variable $z$, we obtain the polynomial $\tilde{P}(z_0) = z_0^N + \sum a_j z_0^{N-j} = 0$. The proof is completed. $\square$

**Theorem 6.36.** *There is no nontrivial 2-dimensional quasi-crystallographic group with translation subgroup* $\mathbb{Z}^3$.

**Proof.** We will use the previous theorem. Let $z_0 = e^{i\varphi}$ be a rotation in such a group, and let $P(z_0) = 0$ be a polynomial of degree 3:

$$z_0^3 + mz_0^2 + nz_0 + p = 0, \quad m, n, p \in \mathbb{Z}.$$

Since $p = 1$, we have

$$P(z) = (z - z_0)(z - \bar{z}_0)(z - 1).$$

Hence the polynomial $P(z)/(z-1)$ has integer coefficients. Therefore, $z_0$ is an algebraic integer of degree 2 and $|z_0| = 1$. This implies that $z_0^4 = 1$ or $z_0^6 = 1$. Thus we arrive at the case of ordinary crystallographic groups, which proves the theorem. $\square$

Consider now the important case where the group $G$ has an Abelian (translation) part isomorphic to $\mathbb{Z}^4$. Assume that

$$z_0 + \bar{z}_0 = 2\cos\varphi = -(m + n\sqrt{d})/2, \quad m, n, d \in \mathbb{Z},$$

with $(m^2 - n^2 d)/4 \in \mathbb{Z}$. The polynomial $P(z)$ has the form

$$\begin{aligned} P(z) &= \left(z^2 - \frac{m + n\sqrt{d}}{2} z + 1\right)\left(z^2 - \frac{m - n\sqrt{d}}{2} z + 1\right) \\ &= z^4 - mz^3 + \left(2 + \frac{m^2 - nd}{4}\right) z^2 - mz + 1, \\ P(z_0) &= 0, \qquad P(\bar{z}_0) = P(z_0^{-1}) = 0. \end{aligned}$$

Then the following two cases are possible:

a) $|m - n\sqrt{d}|/2 < 2$, $|m + n\sqrt{d}|/2 < 2$;

b) $|m - n\sqrt{d}|/2 > 2$, $|m + n\sqrt{d}|/2 < 2$.

**Theorem 6.37.** *The case* a) *holds if and only if the number* $z_0 = e^{i\varphi}$ *is a root of one.*

**Corollary 6.8.** *In case* b) *the number* $z_0 = e^{i\varphi}$ *cannot be a root of one. Thus we arrive at irrational rotation angles.*

**Proof.** If $z_0$ is a root of one, then the corresponding irreducible polynomial with integer coefficients is such that all the other roots are also roots of one. Hence for any other root $w_0 = e^{i\psi}$ we also have

$$|w_0 + \bar{w}_0| = |(m - n\sqrt{d})/2| = |2\cos\psi| < 2.$$

If $|w_0 + \bar{w}_0| > 2$, then $w_0$ surely cannot be a root of one; moreover, $z_0 = e^{i\varphi}$ is not a root of one either. $\square$

EXAMPLE. For even numbers $(m, n)$ the condition $(m^2 - n^2 d)/4 \in \mathbb{Z}$ is fulfilled. Let $-m = n = 2$, $d = 2$. We see that

$$|-2 + 2\sqrt{2}| < 4, \qquad |-2 - 2\sqrt{2}| > 4.$$

Thus the rotation here is irrational. We can easily deduce an infinite series of similar cases from the conditions stated above. The reader can find classification results and very interesting examples for $n = 2$ in [**58**].

For instance, in the case where $n = 2$ and $G \cap \mathbb{R}^2 = \mathbb{Z}^4$, which was discussed above, it is not hard to prove the following general statement.

An element $z_0 = e^{i\varphi}$ is a rotation in a quasi-crystallographic group if and only if there exist integers $m, k \in \mathbb{Z}$, $m^2 \geq 4k$, such that

$$A_+ = z_0 + \bar{z}_0 = \frac{m + \sqrt{m^2 - 4k}}{2}, \qquad |z_0 + \bar{z}_0| < 2$$

(for some sign of $\sqrt{m^2 - 4k}$). In this case we have

$$P(z) = (z^2 - A_+ z + 1)(z^2 - A_- z + 1), \qquad A_- = \frac{m - \sqrt{m^2 - 4k}}{2}.$$

If $w_0$ is another root of the polynomial, then

$$w_0 + \bar{w}_0 = A_-.$$

Thus we conclude that the condition

$$|A_+| < 2, \qquad |A_-| < 2$$

implies that $z_0$ is a root of one, $z_0^l = 1$. Then it is easily seen that the only possibilities are

$$l = 5, 8, 10, 12, \infty$$

(apart from the "crystallographic" $l = 4, 6$ and "trivial" $l = 1, 2, 3$).

Whenever the reverse inequalities hold,

$$|A_+| < 2, \qquad |A_-| > 2, \qquad A_\pm = \frac{m \pm \sqrt{m^2 - 4k}}{2},$$

we have $|z_0| = 1$, but $z_0^l \neq 1$ for any $l \neq 1$.

It turns out that all these cases,

$$l = 5, 8, 10, 12, \infty,$$

are compatible with the local rules (see [**58**]. Thus quasi-crystals with rotational symmetry of infinite order inevitably arise from physical considerations as well.

Now we present a construction demonstrating the geometric (rather than number-theoretic) nature of this phenomenon. Consider the space $\mathbb{R}^4$ with basis $(e_1, e_2, e_3, e_4)$ and integer lattice $\sum m_i e_i$, $m_i \in \mathbb{Z}$. Suppose that in this space there is a Minkowski metric of signature $(1,3)$ and an *irrational space-like direction* $\mathbb{R}^2 \subset \mathbb{R}^4$ such that the group $\mathrm{O}(1,3) \cap \mathrm{SL}(4,\mathbb{Z})$ contains a nontrivial transformation preserving the Minkowski metric, the lattice $\mathbb{Z}^4$, and this direction,

$$T\colon \mathbb{R}^4 \to \mathbb{R}^4, \quad T(\mathbb{Z}^4) \subset \mathbb{Z}^4, \quad T(\mathbb{R}^2) \subset \mathbb{R}^2.$$

This immediately gives rise to a quasi-crystallographic group $G_T\colon \mathbb{R}^2 \to \mathbb{R}^2$, which is defined as follows: let $\mathbb{R}^\perp$ be the "orthogonal" subspace,

$$\mathbb{R}^\perp \oplus \mathbb{R}^2 = \mathbb{R}^4, \qquad \langle \mathbb{R}^\perp, \mathbb{R}^2 \rangle = 0.$$

All the basis vectors $e_j$ determine translations $\mathbb{R}^2 \to \mathbb{R}^2$, so that

$$e_j(\eta) = P_\perp(\eta + e_j), \quad j = 1,2,3,4,$$

and $P_\perp\colon \mathbb{R}^4 \to \mathbb{R}^2$ is the projection in the direction of $\mathbb{R}^\perp$. Moreover, we have the rotation

$$T\colon \mathbb{R}^2 \to \mathbb{R}^2.$$

Altogether they specify a quasi-crystallographic group $G_T \subset \mathrm{E}(\mathbb{R}^2)$ on the (Euclidean) plane, where $|G_T/(G_T \cap \mathbb{R}^2)| = \infty$. Thus, following the above scheme, we obtain the corresponding quasi-lattices as well. Obviously, this construction can be extended to all dimensions. Hence we conclude that

*For a description of the symmetry of real crystals in ordinary Euclidean spaces of dimension $n \geq 2$, one must invoke hyperbolic geometry and the corresponding number-theoretical methods.*

EXAMPLE. Here $n = 2$. As above, let

$$A_\pm = \frac{m \pm \sqrt{m^2 - 4k}}{2}, \quad |A_+| < 2, \quad |A_-| > 2.$$

Consider 4 vectors $e_1, e_2, e_3, e_4 \in \mathbb{R}^4$, the integer lattice $\mathbb{Z}^4$, and the subspace $\mathbb{R}^2 \subset \mathbb{R}^4$ spanned by the vectors

$$\left.\begin{aligned} \eta_1 &= e_1 + e_3 - A_- e_2 \\ \eta_2 &= e_2 + e_4 - A_- e_3 \end{aligned}\right\} \in \mathbb{R}^2.$$

The integer-valued operator $T\colon \mathbb{Z}^4 \to \mathbb{Z}^4$ will be specified in the basis $(e_1, e_2, e_3, e_4)$ by the formula

$$T = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & m \\ 0 & 1 & 0 & -2-k \\ 0 & 0 & 1 & m \end{pmatrix}.$$

In the plane $\mathbb{R}^2$ the transformation $T$ acts as the rotation through the angle $\varphi$ (such that $2\cos\varphi = A_+$). In the plane $\mathbb{R}^\perp$ it acts as the hyperbolic rotation with eigenvalues $w_1$, $w_2$ ($w_j \in \mathbb{R}$), which together with $(z_0, \bar{z}_0)$, where $z_0 = e^{i\varphi}$, are the roots of the polynomial

$$P(z) = z^4 - mx^3 + (2+k)z^2 - mx + 1.$$

In this way, the Minkowski metric is naturally introduced, where $T \in \mathrm{O}(1,3)$.

## Exercises to Chapter 6

1. Show that the group $\mathrm{O}(p,q)$ is diffeomorphic to the manifold $\mathrm{O}(p) \times \mathrm{O}(q) \times \mathbb{R}^{pq}$.

2. Show that the projective spaces $\mathbb{R}P^n$, $\mathbb{C}P^n$, and $\mathbb{H}P^n$ are compact.

3. Construct an atlas of $n+1$ charts on the $n$-dimensional real projective space $\mathbb{R}P^n$.

4. Show that the Stiefel and Grassmann manifolds $V_{n,k}$ and $G_{n,k}$ are compact.

5. Show that the quaternion projective space $\mathbb{H}P^n$ parametrizes the classes of nonzero vectors in $\mathbb{H}^{n+1}$ up to left multiplication by nonzero quaternions.

6. Construct diffeomorphisms $\mathbb{C}P^1 \approx S^2$ and $\mathbb{H}P^1 \approx S^4$.

7. Construct the mappings

$$S^{4n+3} \to \mathbb{H}P^n$$

analogous to the mappings $S^{2n+1} \to \mathbb{C}P^n$ and find the inverse images of points under these mappings.

8. Show that the theta-function satisfies the following periodicity conditions:

$$\theta(z + e_\alpha) = \theta(z), \qquad \theta(z + Be_\alpha) = e^{-2\pi i(z_k + B_{kk}/2)}\theta(z),$$

where $e_1, \ldots, e_n$ is a basis in $\mathbb{C}^n$ and $z = \sum_k z_k e_k$, and that it is an even function,

$$\theta(z) = \theta(-z).$$

9. Show that the groups $\mathrm{GL}(n,\mathbb{C})$, $\mathrm{SL}(n,\mathbb{C})$, $\mathrm{U}(n)$, and $\mathrm{Sp}(n)$ are connected.

10. Show that the group $\mathrm{GL}(n,\mathbb{R})$ consists of two connected components.

11. Let $G$ be a connected Lie group. Show that:

a) $G$ is generated by an arbitrarily small neighborhood of the identity;

b) the connected component of the identity is a normal subgroup in $G$.

12. Show that every Lie group is orientable.

13. Find all the one-parameter subgroups in $\mathrm{SL}(2,\mathbb{R})$ and prove that the image of the exponential mapping does not cover the entire group.

14. Prove the following isomorphisms of Lie algebras:

$$\mathfrak{su}(1,1) \simeq \mathfrak{sl}(2,\mathbb{R}), \qquad \mathfrak{so}(1,3) \simeq \mathfrak{sl}(2,\mathbb{C}).$$

15. Show that the Lie algebra $\mathfrak{so}(1,2)$ is isomorphic to the algebra of vectors in $\mathbb{R}^{1,2}$ relative to the "vector product".

16. Show that Killing's metric on a compact matrix Lie group has the form

$$dl^2(g) = -\operatorname{Tr}(g^{-1}\,dg \cdot g^{-1}\,dg).$$

17. For Killing's metric $g_{ij}$ on a Lie algebra, show that the tensor

$$c_{kij} = g_{kl}c^l_{ij}$$

is skew-symmetric in all the three subscripts (here the $c^k_{ij}$ are structural constants, $[e_i, e_j] = c^k_{ij}e_k$).

18. Show that Killing's metric for the group $\mathrm{SO}(p,q)$ is obtained as a restriction of a pseudo-Euclidean metric $\langle X, Y\rangle = \operatorname{Tr}(GXGY^T)$, where $G$ is the Gram matrix of type $(p,q)$.

19. Show that Euler's angles are coordinates of the second type in the Lie group $\mathrm{SO}(3)$.

20. Let $m = (m_1, \ldots, m_k)$ be a partition of a number $n$, i.e.,

$$m_1 + m_2 + \cdots + m_k = n.$$

A family of linear subspaces of $\mathbb{R}^n$,

$$0 = \pi_0 \subset \pi_1 \subset \cdots \subset \pi_k = \mathbb{R}^n,$$

is called an *m-flag* if

$$\dim \pi_i - \dim \pi_{i-1} = m_i.$$

On the set of all $m$-flags $F(n,m)$, introduce the structure of a homogeneous space (*flag space*) of the group $\mathrm{O}(n)$ and find the isotropy group of this space.

Chapter 7

# Tensor Algebra

## 7.1. Tensors of rank 1 and 2

**7.1.1. Tangent space and tensors of rank 1.** If $x^1, \dots, x^n$ are local coordinates in a domain $U$ on a manifold, then a tangent vector $\xi$ is specified by a $(2n)$-tuple of real numbers: the coordinates $(x^1, \dots, x^n)$ of the initial point and the coordinates $(\xi^1, \dots, \xi^n)$ in the space of vectors originating from this point. The tangent space at a point $x$ is denoted by $T_x M$.

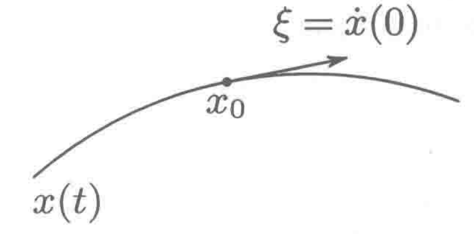


**Figure 7.1.** Tangent vector.

Any such tangent vector can be realized as the velocity vector of a curve on the manifold. Namely, consider the curve written in local coordinates as

$$x(t) = x_0 + \xi t,$$

where $(x_0^1, \dots, x_0^n)$ are the coordinates of the initial point $x$ of the vector $\xi$ (or, in other words, the point where $\xi$ is tangent to the manifold). Then

$$\dot{x}(0) = \frac{dx(t)}{dt}\Big|_{t=0} = \xi.$$

Due to the physical meaning of the velocity vector, it is a definite entity independent of the choice of the coordinate system. However, it is written

differently in different coordinates. Indeed, let us introduce new coordinates

$$z^i = z^i(x^1, \dots, x^n), \quad i = 1, \dots, n,$$

and write the curve $x(t)$ in terms of these new coordinates:

$$z(t) = (z^1(x^1(t), \dots, x^n(t)), \dots, z^n(x^1(t), \dots, x^n(t))).$$

By the chain rule, the velocity vector at $t = 0$ equals

$$\dot{z}(0) = \frac{dz(x(t))}{dt}\Big|_{t=0} = \Big(\frac{\partial z^1}{\partial x^j}\frac{dx^j(t)}{dt}, \dots, \frac{\partial z^n}{\partial x^j}\frac{dx^j(t)}{dt}\Big)\Big|_{t=0}.$$

Here the partial derivatives $\frac{\partial z^i}{\partial x^j}$ are taken at $x = x(0)$ and summation over the repeated subscript and superscript $j$ is assumed. Hence we conclude that the tangent vector to the curve $x(t)$ in coordinates $(z)$ is written as

$$\tilde{\xi}^i = \frac{\partial z^i}{\partial x^j}\,\xi^j.$$

We have already encountered this phenomenon in a simpler setting. Let $e_1, \dots, e_n$ and $\tilde{e}_1, \dots, \tilde{e}_n$ be two bases in the Cartesian space $\mathbb{R}^n$, related by the formula

$$e_j = a^i_j \tilde{e}_i.$$

Then the same vector $\xi$ is written differently in these bases:

$$\xi = \xi^i e_i = \tilde{\xi}^j \tilde{e}_j.$$

We have $\tilde{\xi}^i\tilde{e}_i - \xi^j(a^i_j\tilde{e}_i) = (\tilde{\xi}^i - a^i_j\xi^j)\tilde{e}_i = 0$, which implies the formula relating these two representations:

$$\tilde{\xi}^i = a^i_j\xi^j.$$

Since the change of coordinates is linear, the elements of the matrix $a^i_j = \frac{\partial z^i}{\partial x^j}$ are constant (independent of the point $x$).

Thus we have proved the following theorem.

**Theorem 7.1.** *The velocity vectors $\xi = \dot{x}$ and $\tilde{\xi} = \dot{z}$ of a moving point written in coordinate systems $(x^1, \dots, x^n)$ and $(z^1, \dots, z^n)$ are related by the formula*

$$\tilde{\xi}^i = \frac{\partial z^i}{\partial x^j}\,\xi^j. \tag{7.1}$$

Henceforth we always assume summation over repeated subscripts and superscripts unless otherwise stated.

In mechanics and physics the state of a system is described by a point of an $n$-dimensional space. This space, which comprises all possible configurations of the system, is called the *configuration space*. If the point $x(t)$ moves (the system changes in time), then the state of $x$ is described by the $(2n)$-tuple that consists of its coordinates $(x^1, \dots, x^n)$ and the components of the

velocity vector $\frac{dx}{dt} = (\dot{x}^1, \dots, \dot{x}^n)$. These tuples $(x(t), \dot{x}(t))$ are specified by points of the $2n$-dimensional *phase space.*

In applications it may be expedient to use different coordinate systems. For example, we already encountered cylindrical and spherical coordinates in $\mathbb{R}^3$, which are related to affine coordinates by nonlinear transformations of the form

$$z^i = z^i(x^1, \dots, x^n), \quad i = 1, \dots, n.$$

Furthermore, if the configuration space has a complicated geometry (e.g., it is a two-dimensional sphere), then it is impossible to introduce a single coordinate system on the entire space, and we must cover it by charts with coordinates related by transition formulas on the intersections of charts. In any case, the velocity vectors $\dot{z}$ and $\dot{x}$ in different coordinates are related by formula (7.1).

The quantities which remain unchanged under coordinate changes are called *scalars.* Real-valued functions $f(x^1, \dots, x^n)$ are an example. Theorem 7.1 shows that such functions are insufficient for describing physical systems of interest. We need a new type of quantities, which are called *tensors.*

If the expression for $(\xi^1, \dots, \xi^n)$ changes under a change of coordinates by formula (7.1), it is called a *vector.* A vector is the simplest example of a tensor. We now give other natural examples of tensors.

The expression

$$\operatorname{grad} f = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}\right)$$

determines the *gradient of a function.*

**Theorem 7.2.** *The value of the gradient depends on the choice of the coordinate system. Under a change of coordinates the quantities*

$$\xi = (\xi_1, \dots, \xi_n) = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}\right) \text{ and } \tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_n) = \left(\frac{\partial f}{\partial z^1}, \dots, \frac{\partial f}{\partial z^n}\right)$$

*are related by the formula*

$$\tilde{\xi}_i = \frac{\partial x^j}{\partial z^i} \xi_j. \tag{7.2}$$

**Proof.** By the chain rule, for a composite function $f(x) = f(z(x))$, we have

$$\frac{\partial f}{\partial z^i} = \frac{\partial f}{\partial x^j} \frac{\partial x^j}{\partial z^i}.$$

The proof is completed. □

Observe that formula (7.1) is different from (7.2), and so the gradient of a function is not a vector. It exemplifies another type of tensor: a quantity $(\xi_1, \dots, \xi_n)$, which under the change of coordinates transforms by formula (7.2), is called a *covector*.

Recall that the Jacobi matrix $J$ of a mapping (coordinate change) $(x) \to (z)$ is the matrix

$$J = (a^i_j) = \left(\frac{\partial z^i}{\partial x^j}\right).$$

Its transpose is $J^\top = (b^i_j)$, where $a^i_j = b^j_i$. Thus formulas (7.1) and (7.2) become

$$\tilde{\xi} = J\xi \quad \text{(velocity vector)}, \qquad \xi = J^\top \tilde{\xi} \quad \text{(gradient)}.$$

The changes of coordinates are invertible, hence we can rewrite the formula for gradient as

$$\tilde{\xi} = (J^\top)^{-1}\xi.$$

We come to an important conclusion: the vectors and covectors are transformed in the same manner if $J = (J^\top)^{-1}$. This condition means that $J J^\top = 1$, i.e., the Jacobi matrix specifies an orthogonal transformation. Therefore, in the case of orthonormal coordinates in Euclidean space, we sometimes speak about a gradient as a vector and do not distinguish between subscripts and superscripts because the vectors and covectors are transformed identically when changing to other orthonormal coordinates.

**Lemma 7.1.** *Let $(z) \to (x)$ be the change of coordinates inverse to the change $(x) \to (z)$. Then*

$$\frac{\partial x^i}{\partial z^j}\frac{\partial z^j}{\partial x^k} = \delta^i_k. \tag{7.3}$$

**Proof.** We apply the chain rule to the identity mapping $x = x(z(x))$. Then the left- and right-hand sides of (7.3) are equal to $\frac{\partial x^i}{\partial x^k}$, which proves the lemma. □

This lemma implies the following important theorem.

**Theorem 7.3.** *Covectors are linear functions of vectors: if $\eta = (\eta_1, \dots, \eta_n)$ is a covector at a point $x$, then it specifies a linear function $\eta\colon V \to \mathbb{R}$ on the space $V$ of vectors at this point by the formula*

$$\eta(\xi) = \eta_i \xi^i.$$

**Proof.** Linearity of this function is obvious. It remains to prove that its value does not depend on the choice of coordinates. When changing to

new coordinates $z = z(x)$, we have, by formulas (7.1) and (7.2), $\tilde{\xi}^i = \frac{\partial z^i}{\partial x^j}\,\xi^j$ and $\tilde{\eta}_i = \frac{\partial x^k}{\partial z^i}\,\eta_k$, whence

$$\tilde{\eta}_i\tilde{\xi}^i = \Big(\frac{\partial x^k}{\partial z^i}\,\eta_k\Big)\Big(\frac{\partial z^i}{\partial x^j}\,\xi^j\Big) = \delta^k_j\eta_k\xi^j = \eta_k\xi^k.$$

The proof is completed. □

If a covector $\eta = \operatorname{grad} f$ is the gradient of a function $f$, then the corresponding linear function

$$\eta(\xi) = \frac{\partial f}{\partial x^i}\,\xi^i$$

is known in calculus as the directional derivative of $f$ in the direction of $\xi$.

The scalars have no indices, hence they are said to be tensors of rank 0. The vectors and covectors involve one index (a subscript or superscript), so they are tensors of rank 1. The vectors are said to be of type $(1,0)$ (one superscript), and covectors are of type $(0,1)$ (one subscript).

**7.1.2. Tensors of rank 2.** An example of a tensor of rank 2 is given by the Riemannian metric $g_{ij}(x)$.

**Theorem 7.4.** *Under coordinate changes $z = z(x)$ the Riemannian metric $g_{ij}\,dx^i\,dx^j = \tilde{g}_{kl}\,dz^k\,dz^l$ is transformed by the formula*

$$\tilde{g}_{kl} = g_{ij}\,\frac{\partial x^i}{\partial z^k}\,\frac{\partial x^j}{\partial z^l}. \tag{7.4}$$

**Proof.** The Riemannian metric assigns to each pair of vectors $\xi$ and $\eta$ that are tangent at a point $x$ their scalar product

$$\langle \xi, \eta\rangle = g_{ij}(x)\xi^i\eta^j.$$

Since the length of the curve $x(t)$, $a \le t \le b$, equals $\int_a^b \sqrt{\langle \dot{x}, \dot{x}\rangle}\,dt$ and does not depend on the choice of coordinates, the scalar product $\langle \xi, \eta\rangle$ does not depend on the choice of coordinates either. Under the change of coordinates we have $\xi^i = \frac{\partial x^i}{\partial z^k}\,\tilde{\xi}^k$ and $\eta^j = \frac{\partial x^j}{\partial z^l}\,\tilde{\eta}^l$. Hence

$$g_{ij}\xi^i\eta^j = g_{ij}\,\frac{\partial x^i}{\partial z^k}\,\tilde{\xi}^k\,\frac{\partial x^j}{\partial z^l}\,\tilde{\eta}^l = \Big(g_{ij}\,\frac{\partial x^i}{\partial z^k}\,\frac{\partial x^j}{\partial z^l}\Big)\tilde{\xi}^k\tilde{\eta}^l = \tilde{g}_{kl}\tilde{\xi}^k\tilde{\eta}^l.$$

Since $\xi$ and $\eta$ are arbitrary vectors, the last equality implies (7.4). □

Covectors form a linear space relative to the ordinary coordinatewise addition,

$$\xi + \eta = (\xi_1 + \eta_1, \ldots, \xi_n + \eta_n).$$

Moreover, we can define the scalar product for covectors by the formula

$$\langle \xi, \eta\rangle = g^{ij}\xi_j\eta_i.$$

An example of a skew-symmetric scalar product is given by the Poisson structure. In this case $g^{ij} = \{x^i, x^j\}$, where $\{\cdot, \cdot\}$ is the Poisson bracket in the domain. By the definition of this bracket,

$$\tilde{g}^{ij} = \{z^i, z^j\} = \frac{\partial z^i}{\partial x^k}\frac{\partial z^j}{\partial x^l}\{x^k, x^l\} = \frac{\partial z^i}{\partial x^k}\frac{\partial z^j}{\partial x^l}g^{kl}.$$

Thus we obtain the transformation formula for the symbols $g^{ij}$ that specifies the Poisson structure:

$$\tilde{g}^{kl} = g^{ij}\frac{\partial z^k}{\partial x^i}\frac{\partial z^l}{\partial x^j} \tag{7.5}$$

(note that this formula differs from the transformation formula of the Riemannian metric).

The linear operators acting on vectors are specified by matrices with entries having one subscript and one superscript:

$$\xi^i = a^i_j\eta^j, \quad \xi = A\eta.$$

Under the coordinate changes $\tilde{\xi}^i = \frac{\partial z^i}{\partial x^k}\xi^k$, $\tilde{\eta}^j = \frac{\partial z^j}{\partial x^l}\eta^l$ the equality $\tilde{\xi}^i = \tilde{a}^i_j\tilde{\eta}^j$ becomes

$$\Big(\frac{\partial z^i}{\partial x^k}\xi^k\Big) = \tilde{a}^i_j\Big(\frac{\partial z^j}{\partial x^l}\eta^l\Big).$$

Multiplying both sides by $\frac{\partial x^m}{\partial z^i}$ and summing over $i$, we obtain, taking Lemma 7.1 into account,

$$\xi^m = \Big(\tilde{a}^i_j\frac{\partial z^j}{\partial x^l}\frac{\partial x^m}{\partial z^i}\Big)\eta^l = a^m_l\eta^l.$$

This implies the following result.

**Theorem 7.5.** *The elements of the matrix $A = (a^i_j)$ of a linear operator are transformed under coordinate changes by the rule*

$$\tilde{a}^i_j = \frac{\partial z^i}{\partial x^k}\frac{\partial x^l}{\partial z^j}a^k_l. \tag{7.6}$$

Formulas (7.4)–(7.6) specify the transformation rules for all tensors of rank 2, although we derived them in special settings (e.g., tensors specifying the Riemannian metric are always symmetric, $g_{ij} = g_{ji}$, whereas those specifying the Poisson structures are skew-symmetric, $g^{ij} = -g^{ji}$).

**7.1.3. Transformations of tensors of rank at most 2.** Here we give a list of transformation rules $T \to \tilde{T}$ derived so far for tensors of rank at most 2 under the coordinate changes

$$(x^1, \dots, x^n) \to (z^1, \dots, z^n), \qquad z^i = z^i(x^1, \dots, x^n), \quad i = 1, \dots, n.$$

1. Tensors of rank 0 (scalars) do not change:

$$\tilde{T} = T.$$

2. Tensors of rank 1:

2a) vectors (type $(1,0)$, e.g., the velocity vectors):

$$\tilde{T}^i = \frac{\partial z^i}{\partial x^j} T^j;$$

2b) covectors (type $(0,1)$, e.g., the gradients of functions):

$$\tilde{T}_i = \frac{\partial x^j}{\partial z^i} T_j.$$

3. Tensors of rank 2:

3a) tensors of type $(0,2)$ (e.g., Riemannian metrics):

$$\tilde{T}_{ij} = \frac{\partial x^k}{\partial z^i} \frac{\partial x^l}{\partial z^j} T_{kl};$$

3b) tensors of type $(2,0)$ (e.g., Poisson structures):

$$\tilde{T}^{ij} = \frac{\partial z^i}{\partial x^k} \frac{\partial z^j}{\partial x^l} T^{kl};$$

3c) linear operators on vectors (type $(1,1)$):

$$\tilde{T}^i_j = \frac{\partial z^i}{\partial x^k} \frac{\partial x^l}{\partial z^j} T^k_l.$$

## 7.2. Tensors of arbitrary rank

**7.2.1. Transformation of components.** The above transformation rules under coordinate changes for tensors of small rank lead us to the general definition of tensor.

A *tensor* (or a *tensor field*) is an object specified in every coordinate system $(x^1, \dots, x^n)$ by an array of numbers $T^{i_1 \dots i_p}_{j_1 \dots j_q}$ which transform under a coordinate change $(x) \to (z)$: $x^i = x^i(z^1, \dots, z^n)$, $z^j = z^j(x^1, \dots, x^n)$, $i, j = 1, \dots, n$, $z(x(z)) = z$, by the rule

$$\tilde{T}^{i_1 \dots i_p}_{j_1 \dots j_q} = \frac{\partial z^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial z^{i_p}}{\partial x^{k_p}} \frac{\partial x^{l_1}}{\partial z^{j_1}} \cdots \frac{\partial x^{l_q}}{\partial z^{j_q}} T^{k_1 \dots k_p}_{l_1 \dots l_q}, \tag{7.7}$$

where $\tilde{T}^{i_1 \dots i_p}_{j_1 \dots j_q}$ is the tensor written in coordinates $(z)$ and $T^{k_1 \dots k_p}_{l_1 \dots l_q}$ is the same tensor in coordinates $(x)$. A tensor $T$ with $p$ superscripts and $q$ subscripts is said to be of *type* $(p,q)$ and *rank* $p+q$.

Recall that the superscripts and subscripts in (7.7) run over $1, \dots, n$, where $n$ is the dimension of the space, and summation over repeated superscripts and subscripts is assumed.

We can deduce from (7.7) the formula for the inverse transformation. To this end, multiply both sides of (7.7) by $\frac{\partial x^{r_1}}{\partial z^{i_1}} \cdots \frac{\partial x^{r_p}}{\partial z^{i_p}} \frac{\partial z^{j_1}}{\partial x^{s_1}} \cdots \frac{\partial z^{j_q}}{\partial x^{s_q}}$ and sum

over repeated indices $i_1, \ldots, i_p, j_1, \ldots, j_q$. Taking into account Lemma 7.1, which states that $\frac{\partial x^i}{\partial z^j}\frac{\partial z^j}{\partial x^k} = \delta^i_k$, we obtain

$$\begin{aligned}
&\frac{\partial x^{r_1}}{\partial z^{i_1}} \cdots \frac{\partial x^{r_p}}{\partial z^{i_p}} \frac{\partial z^{j_1}}{\partial x^{s_1}} \cdots \frac{\partial z^{j_q}}{\partial x^{s_q}} \tilde{T}^{i_1 \ldots i_p}_{j_1 \ldots j_q} \\
&\quad = \frac{\partial x^{r_1}}{\partial z^{i_1}} \cdots \frac{\partial x^{r_p}}{\partial z^{i_p}} \frac{\partial z^{j_1}}{\partial x^{s_1}} \cdots \frac{\partial z^{j_q}}{\partial x^{s_q}} \frac{\partial z^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial z^{i_p}}{\partial x^{k_p}} \frac{\partial x^{l_1}}{\partial z^{j_1}} \cdots \frac{\partial x^{l_q}}{\partial z^{j_q}} T^{k_1 \ldots k_p}_{l_1 \ldots l_q} \\
&\quad = \delta^{r_1}_{k_1} \cdots \delta^{r_p}_{k_p} \delta^{l_1}_{s_1} \cdots \delta^{l_q}_{s_q} T^{k_1 \ldots k_p}_{l_1 \ldots l_q} = T^{r_1 \ldots r_p}_{s_1 \ldots s_q}.
\end{aligned}$$

As a result, we obtain the following inversion formula for transformation (7.7):

$$T^{r_1 \ldots r_p}_{s_1 \ldots s_q} = \frac{\partial x^{r_1}}{\partial z^{i_1}} \cdots \frac{\partial x^{r_p}}{\partial z^{i_p}} \frac{\partial z^{j_1}}{\partial x^{s_1}} \cdots \frac{\partial z^{j_q}}{\partial x^{s_q}} \tilde{T}^{i_1 \ldots i_p}_{j_1 \ldots j_q}.$$

This formula determines the inverse transformation of tensors under the inverse transformation of coordinates $(z) \to (x)$ and is obtained from (7.7) by replacement $(x) \leftrightarrow (z)$. The above calculation may be regarded as a verification of the fact that the transition formulas do not depend on the choice of coordinates $(x)$ or $(z)$.

**Theorem 7.6.** *At any point of the $n$-dimensional space, the tensors of type $(p, q)$ form a linear space relative to the operations of addition*

$$(T + S)^{i_1 \ldots i_p}_{j_1 \ldots j_q} = T^{i_1 \ldots i_p}_{j_1 \ldots j_q} + S^{i_1 \ldots i_p}_{j_1 \ldots j_q}$$

*and multiplication by scalars*

$$(\lambda T)^{i_1 \ldots i_p}_{j_1 \ldots j_q} = \lambda T^{i_1 \ldots i_p}_{j_1 \ldots j_q}.$$

*The dimension of this space is $n^{p+q}$.*

**Proof.** The fact that tensors form a linear space follows from (7.7). By this formula, the expressions $T + S$ and $\lambda T$ also transform as tensors.

Let $V = \mathbb{R}^n$ be the space of vectors at a given point and let $e_1, \ldots, e_n$ be its basis. The covectors constitute the *dual* vector space $V^*$, the dual basis of which, $e^1, \ldots, e^n$, is determined by the formulas

$$e^i(e_j) = \delta^i_j, \quad i, j = 1, \ldots, n.$$

Indeed, covectors are linear functions of vectors (Theorem 7.3), and the value of any linear function $\xi$ on the vector $\eta = \eta^i e_i$ is equal to $\xi(\eta) = \eta^i \xi(e_i)$. The expansion of the function $\xi$ in the dual basis has the form

$$\xi = \xi_i e^i, \quad \xi_i = \xi(e_i).$$

Denote by

$$e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes \cdots \otimes e^{j_q}, \tag{7.8}$$

where $1 \le i_1, \ldots, i_p, j_1, \ldots, j_q \le n$, the tensor $S$ of type $(p, q)$ with all components equal to zero, except for one component $S^{i_1 \ldots i_p}_{j_1 \ldots j_q} = 1$. Such tensors form a basis in the space of tensors of type $(p, q)$ at a given point, since any tensor $T$ is uniquely representable as a linear combination

$$T = T^{i_1 \ldots i_p}_{j_1 \ldots j_q} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes \cdots \otimes e^{j_q}.$$

The operations of addition and multiplication by a scalar in this formula are defined coordinatewise.

The order of indices $i_1, \ldots, i_p$ and $j_1, \ldots, j_q$ in (7.8) is essential: their permutation yields a different basis tensor. Therefore, this basis contains $n^{p+q}$ elements. □

Note that the addition operation is defined only for tensors "attached" to one and the same point.

**7.2.2. Algebraic operations on tensors.** Here we define the basic algebraic operations on tensors.

In order to introduce the operation of index permutation, recall the definition of the permutation group $S_n$ (also called the *symmetric group*).

Consider the integers from 1 to $n$ and reorder them in some way. Such a permutation is determined by a one-to-one mapping $\sigma\colon X \to X$ of the set $X = \{1, \ldots, n\}$ onto itself:

$$\sigma = \begin{pmatrix} 1 & & n \\ \sigma(1) & & \sigma(n) \end{pmatrix}.$$

The symmetric group $S_n$ contains all such permutations with composition as the group operation. The unit element of this group is the identity permutation, $\sigma(k) = k$, $k = 1, \ldots, n$.

Each permutation $\sigma$ may be obtained as composition of transpositions of two elements:

$$\begin{pmatrix} 1 & & k & & l & & n \\ 1 & \ldots & l & & k & \ldots & n \end{pmatrix}.$$

A permutation $\sigma$ is said to be even, $\operatorname{sgn}(\sigma) = 1$, if it is representable as the composition of an even number of transpositions. Otherwise the permutation is called odd, $\operatorname{sgn}(\sigma) = -1$.

Let $A_\sigma\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping defined by the formula $A_\sigma(e_i) = e_{\sigma(i)}$, $i = 1, \ldots, n$. We will denote the matrix specifying this mapping in the

basis $e_1, \ldots, e_n$ also by $A_\sigma$, and we have $\det A_\sigma = \operatorname{sgn}(\sigma)$. This representation implies that

1) The group of permutations is a subgroup of the group $\mathrm{O}(n) \subset \mathrm{GL}(n)$; the mapping

$$\sigma \to A_\sigma \tag{7.9}$$

determines an imbedding of $S_n$ into $\mathrm{O}(n)$.

2) The mapping

$$\operatorname{sgn}\colon S_n \to \{\pm 1\}$$

is a homomorphism, and its kernel is the group of even permutations $A_n \subset S_n$.

3) Under the mapping (7.9) the group of even permutations is imbedded into $\mathrm{SO}(n)$.

For each permutation $\sigma$ of the form $\sigma(k) = i_k$, $k = 1, \ldots, n$, we define the symbol

$$\varepsilon_{i_1 \ldots i_n} = \begin{cases} 1 & \text{if } \sigma \text{ is even}, \\ -1 & \text{if } \sigma \text{ is odd}. \end{cases}$$

For $n = 3$, this symbol was given by formula (6.19).

Now we introduce algebraic operations on tensors.

1. Permutation of indices. Let $T = T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ be a tensor of type $(p, q)$, and $\sigma \in S_q$ a permutation of the numbers $(1, \ldots, q)$.

Define the tensor $\sigma(T)$ by formula

$$\sigma(T)^{i_1 \ldots i_p}_{j_1 \ldots j_q} = T^{i_1 \ldots i_p}_{j_{\sigma(1)} \ldots j_{\sigma(q)}}.$$

This tensor is obtained from $T$ by the permutation of subscripts specified by $\sigma$. Permutation of superscripts is defined in a similar way. It can be easily deduced from (7.7) that the operations of permutation of superscripts and subscripts are not invariant relative to coordinate changes.

2. Contraction (trace). For any pair consisting of a superscript and a subscript we can produce the *contraction* of the tensor. This operation results in the *trace* of the tensor. It is defined as follows.

Let $T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$ be a tensor of type $(p, q)$, and $(i_k, j_l)$ a pair of indices. The tensor $S$ of type $(p-1, q-1)$ is defined as

$$S^{i_1 \ldots i_{p-1}}_{j_1 \ldots j_{q-1}} = T^{i_1 \ldots i_{k-1} i i_k \ldots i_{p-1}}_{j_1 \ldots j_{l-1} i j_l \ldots j_{q-1}}$$

and is called the trace of $T$. Recall that in this formula summation over the repeated index $i$ is assumed.

In the simplest case where $T$ is the tensor specifying a linear operator, we already encountered this operation, which associates with the matrix $T^i_j$ its trace

$$S = \operatorname{Tr} T = T^i_i.$$

3. TENSOR MULTIPLICATION (PRODUCT). Let $T$ and $S$ be tensors of types $(p,q)$ and $(k,l)$. The *tensor product* of $T$ and $S$ is defined as the tensor $R = T \otimes S$ of type $(p+k, q+l)$ given by the formula

$$R^{i_1 \dots i_{p+k}}_{j_1 \dots j_{q+l}} = T^{i_1 \dots i_p}_{j_1 \dots j_q} S^{i_{p+1} \dots i_{p+k}}_{j_{q+1} \dots j_{q+l}}.$$

**Theorem 7.7.** 1. *The operations of permutation of indices, contraction, and tensor multiplication result again in tensors.*

2. *The operation of permutation of indices is linear:*

$$\sigma(T+S) = \sigma(T) + \sigma(T), \qquad \sigma(\lambda T) = \lambda \sigma(T).$$

3. *Tensor multiplication is linear, associative, but not commutative:*

$$\begin{gathered}(T_1 + T_2) \otimes S = T_1 \otimes S + T_2 \otimes S, \quad (\lambda T) \otimes S = \lambda (T \otimes S), \\ T \otimes (S_1 + S_2) = T \otimes S_1 + T \otimes S_2, \quad T \otimes (\lambda S) = \lambda (T \otimes S), \\ T \otimes (S \otimes R) = (T \otimes S) \otimes R.\end{gathered}$$

**Proof.** Statements 2 and 3 follow directly from definitions. Only statement 1 needs a detailed proof.

We will denote by $T$ and $S$ the tensors written in coordinates $(x)$, and by $\tilde{T}$ and $\tilde{S}$ the same tensors in coordinates $(z)$.

a) Since any permutation is a product of transpositions, it suffices to prove that interchanging a single pair of indices results in a tensor. Let

$$\sigma = \begin{pmatrix} 1 & & r & \dots & s & \dots & q \\ 1 & & s & \dots & r & & q \end{pmatrix}.$$

Then the tensor $S = \sigma(T)$ has the form

$$S^{i_1 \dots i_p}_{j_1 \dots j_r \dots j_s \dots j_q} = T^{i_1 \dots i_p}_{j_1 \dots j_s \dots j_r \dots j_q}.$$

On changing to coordinates $(z^1, \dots, z^n)$ we obtain

$$\begin{aligned} S^{i_1 \dots i_p}_{j_1 \dots j_q} &= T^{i_1 i_p}_{j_1 \dots j_s \dots j_r \dots j_q} \\ &= \tilde{T}^{k_1 \dots k_p}_{l_1 \dots l_q} \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \frac{\partial z^{l_r}}{\partial x^{j_s}} \cdots \frac{\partial z^{l_s}}{\partial x^{j_r}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}}. \end{aligned} \tag{7.10}$$

Because of summation over $l_r$ and $l_s$, their interchanging does not affect the right-hand side of (7.10), which becomes

$$S^{i_1\dots i_p}_{j_1\dots j_q} = \tilde{T}^{k_1\dots k_p}_{l_1\dots l_s\dots l_r\dots l_q} \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \frac{\partial z^{l_s}}{\partial x^{j_s}} \cdots \frac{\partial z^{l_r}}{\partial x^{j_r}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}}$$
$$= \tilde{S}^{k_1\dots k_p}_{l_1\dots l_q} \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \frac{\partial z^{l_r}}{\partial x^{j_r}} \cdots \frac{\partial z^{l_s}}{\partial x^{j_s}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}}.$$

Hence $\sigma(T) = S$ is a tensor of type $(p, q)$.

b) Let $S$ be the tensor obtained from $T$ by contraction over $i_r$ and $j_s$:

$$S^{i_1\dots\widehat{i_r}\dots i_p}_{j_1\dots\widehat{j_s}\dots j_q} = \tilde{T}^{k_1\dots k_p}_{l_1\dots l_q} \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}}\Big|_{i_r=j_s=i}$$
$$= \tilde{T}^{k_1\dots k_p}_{l_1\dots l_q} \delta^{l_s}_{k_r} \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \widehat{\frac{\partial x^{i_r}}{\partial z^{k_r}}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \widehat{\frac{\partial z^{l_s}}{\partial x^{j_s}}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}}$$
$$= \tilde{S}^{k_1\dots\widehat{k_r}\dots k_p}_{l_1\dots\widehat{l_s}\dots l_q} \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \widehat{\frac{\partial x^{i_r}}{\partial z^{k_r}}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \widehat{\frac{\partial z^{l_s}}{\partial x^{j_s}}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}}.$$

In this formula the notation $\hat{a}$ means that the expression $a$ is skipped. Moreover, we used the equality $\frac{\partial x^i}{\partial z^{k_r}} \frac{\partial z^{l_s}}{\partial x^i} = \delta^{l_s}_{k_r}$. Thus we have proved that the contraction of a tensor $T$ over the indices $i_r$ and $j_s$ is also a tensor.

c) It is obvious that the tensor product of tensors is again a tensor. □

In the space of tensors of type $(p, q)$ at a given point, we introduced the basis (7.8) of the form $e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes \cdots \otimes e^{j_q}$. All these basis tensors can be obtained as tensor products of vectors and covectors. Therefore, the space of tensors (at a point $x$) of type $(p, q)$ is a *tensor product* of linear spaces and has the form

$$V^{p,q} = \underbrace{V \otimes \cdots \otimes V}_{p} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q},$$

where $V$ is the space of vectors (at the point $x$) and $V^*$ is the space of covectors dual to $V$. This form means that the space is linearly generated by tensor products of the basis tensors of their factors. Recall that $\dim V^{p,q} = n^{p+q}$, where $n = \dim V$.

Permutations of superscripts and subscripts of the basis vectors can be linearly extended to the entire space of tensors. This determines the homomorphism of groups of permutations in $\mathrm{GL}(n^{p+q})$:

$$S_p \to \mathrm{GL}(n^{p+q}), \quad S_q \to \mathrm{GL}(n^{p+q}).$$

**7.2.3. Differential notation for tensors.** The transformation (7.7) of a tensor field under a change of coordinates may be regarded as a change of

basis,

$$\tilde{e}_{k_1} \otimes \cdots \otimes \tilde{e}_{k_p} \otimes \tilde{e}^{l_1} \otimes \cdots \otimes \tilde{e}^{l_q}$$
$$= \frac{\partial x^{i_1}}{\partial z^{k_1}} \cdots \frac{\partial x^{i_p}}{\partial z^{k_p}} \frac{\partial z^{l_1}}{\partial x^{j_1}} \cdots \frac{\partial z^{l_q}}{\partial x^{j_q}} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes \cdots \otimes e^{j_q}$$

in the space of tensors. Observe that we know from calculus two other expressions that change by the same rule, namely, the operators of partial differentiation with respect to one of the variables $x^1, \ldots, x^n$ and the differentials $dx^1, \ldots, dx^n$.

Indeed, by the chain rule, for any smooth function $f$,

$$\frac{\partial f}{\partial z^k} = \frac{\partial x^i}{\partial z^k} \frac{\partial f}{\partial x^i},$$

and hence

$$\frac{\partial}{\partial z^k} = \frac{\partial x^i}{\partial z^k} \frac{\partial}{\partial x^i}.$$

Thus we see that the operator of partial differentiation with respect to $x^i$ behaves under coordinate changes as the basis vector $e_i$ in the tangent space. Therefore, we may identify them and write the vector fields as

$$\xi(x) = \xi^i(x) \frac{\partial}{\partial x^i}.$$

The differentials $dx^j$ are transformed by the rule

$$dz^l = dz^l(x^1, \ldots, x^n) = \frac{\partial z^l}{\partial x^j} dx^j,$$

i.e., as the basis covectors $e^j$. Therefore, a covector field is written in the basis $e^1 = dx^1, \ldots, e^n = dx^n$ as

$$\eta(x) = \eta_i(x) \, dx^i.$$

The simplest example of a covector is the gradient $\operatorname{grad} f$ of a function $f$. We know that covectors are linear functions on vector spaces acting by the rule

$$\eta(\xi) = \eta_i \xi^i.$$

Let $\eta = \operatorname{grad} f$ and let $\xi$ be a vector field. Then

$$\operatorname{grad} f(\xi) = \frac{\partial f}{\partial x^i} \xi^i$$

at any point $x$ is the directional derivative of $f$ at this point in the direction of the vector $\xi(x)$.

Another example of a tensor written in the differential form is the Riemannian metric $g_{ij} \, dx^i \, dx^j$. This is nothing but the expansion of a metric tensor of type $(0, 2)$ in the basis $e^i \otimes e^j = dx^i \otimes dx^j$.

**7.2.4. Invariant tensors.** Under a change of coordinates, tensors transform by formula (7.7). A tensor is said to be invariant if it has the same form in all coordinate systems. This means that

$$\tilde{T}^{i_1\ldots i_p}_{j_1\ldots j_q} = T^{i_1\ldots i_p}_{j_1\ldots j_q}$$

for any change of coordinates $(x) \to (z)$.

Let $\langle \xi, \eta \rangle$ be a scalar product in the vector space $V$, and let $G$ be the matrix group of linear transformations preserving this scalar product. For example, this may be $\mathrm{O}(n)$ or $\mathrm{O}(p,q)$. In this case we may consider a weaker invariance property: a tensor is said to be invariant relative to the transformations in $G$ if its components do not change under the coordinate changes with Jacobi matrices in $G$. Then the tensor is invariant if it is invariant relative to transformations in the group $G = \mathrm{GL}(n)$.

A tensor that is invariant relative to orthogonal transformations of $V$ is said to be *isotropic*.

**Theorem 7.8.** 1. *There are no nonzero isotropic tensors of rank* 1.

2. *Among the tensors of type* $(0,2)$ *only those of the form* $\lambda\delta_{ij}$ *are isotropic.*

3. *There are no nonzero isotropic tensors of rank* 3.

4. *The isotropic tensors of type* $(0,4)$ *form a three-parameter family*

$$T_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\delta_{ik}\delta_{jl} + \nu\delta_{il}\delta_{jk}.$$

**Proof.** Statement 1 is obvious because it says that there are no vectors and covectors which remain unchanged under rotations.

We prove statement 2. Suppose the tensor $T_{ij}$ is isotropic. The transformation $z^i = -x^i$, $z^j = x^j$, $i \neq j$, changes only the component $T_{ij}$: $\tilde{T}_{ij} = -T_{ij}$. Hence an isotropic tensor must be of the form $\sum_i \lambda_i \delta_{ij}$. Under the change $z^i = x^j$, $z^j = x^i$, $z^k = x^k$, where $k \neq i, j$, we have $\tilde{T}_{ii} = T_{jj}$, $\tilde{T}_{jj} = T_{ii}$. Therefore, all the constants $\lambda_i$ coincide, and we conclude that an isotropic tensor must have the form $T_{ij} = \lambda\delta_{ij}$.

It remains to prove that the tensor $\delta_{ij}$ is isotropic. But the matrix $A = \left(\frac{\partial x^i}{\partial z^j}\right)$ is orthogonal if and only if $A^{-1} = A^\top$, and this condition is written as

$$\delta_{ij} = \frac{\partial x^k}{\partial z^i}\frac{\partial x^l}{\partial z^j}\delta_{kl}.$$

Therefore, the tensor $\delta_{ij}$ is isotropic, which proves statement 2.

The proofs of statements 3 and 4 are similar to those of 1 and 2, but are more tedious. □

**Lemma 7.2.** *If $T$ is an invariant tensor, then its type is of the form $(p,p)$ and, in particular, the tensor has an even rank.*

**Proof.** Under a dilatation $z = \lambda x$ the tensor $T$ of type $(p, q)$ transforms into $\lambda^{p-q}T$. Hence it may be invariant only if $p = q$. □

The invariant tensors of small dimension are described by the following analog of Theorem 7.8.

**Theorem 7.9.** 1. *Invariant tensors of rank* 2 *have the form* $T^i_j = \lambda\delta^i_j$.

2. *Invariant tensors of rank* 4 *compose a two-parameter family*

$$T^{ij}_{kl} = \lambda\delta^i_k\delta^j_l + \mu\delta^i_l\delta^j_k.$$

**7.2.5. A mechanical example: strain and stress tensors.** Suppose we have a domain in Euclidean space filled with some continuum, which may be a solid body or a fluid. Consider a situation where exterior forces deform the continuum, which was initially in the equilibrium state.

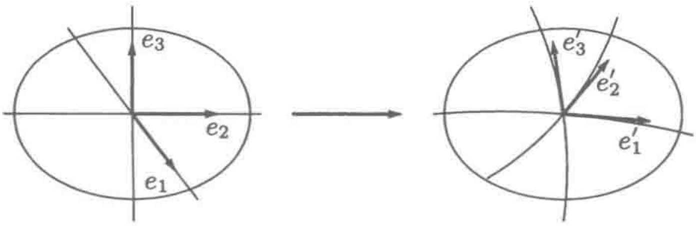


**Figure 7.2.** Deformation of a continuum.

In mathematical terms this deformation is described by the *displacement vector* $u^i = x'^i - x^i$, which specifies the displacement of the point with initial coordinates $(x^1, x^2, x^3)$ into the point $(x'^1, x'^2, x'^3)$. The coordinates $x'^i$ are functions of the coordinates $x^i$. Hence the displacement vector is also a function of the initial coordinates of the point, $u^i = u^i(x^1, x^2, x^3)$, $i = 1, 2, 3$.

The deformation is nontrivial if the distances between points change, which gives rise to forces tending to restore the initial equilibrium state. Since in the new coordinates $x'^i$ the distance between points of the deformed body is also measured in the Euclidean metric

$$ds'^2 = (dx'^1)^2 + (dx'^2)^2 + (dx'^3)^2,$$

we have

$$\begin{aligned} ds'^2 &= (dx^1 + du^1)^2 + (dx^2 + du^2)^2 + (dx^3 + du^3)^2 \\ &= \left(dx^1 + \frac{\partial u^1}{\partial x^i}dx^i\right)^2 + \left(dx^2 + \frac{\partial u^2}{\partial x^i}\, dx^i\right)^2 + \left(dx^3 + \frac{\partial u^3}{\partial x^i}\, dx^i\right)^2 \\ &= (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + 2\sum_{k=1}^{3} \frac{\partial u^k}{\partial x^i}\, dx^k\, dx^i + \sum_{k=1}^{3} \frac{\partial u^k}{\partial x^i}\frac{\partial u^k}{\partial x^j}\, dx^i\, dx^j. \end{aligned}$$

This expression is written in Euclidean coordinates, and, as is usually done in physics and mechanics (and will be justified below), we will not distinguish between superscripts and subscripts. Then the formula for $ds'^2$ is rewritten as

$$ds'^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + \Big(\frac{\partial u_i}{\partial x^k} + \frac{\partial u_k}{\partial x^i}\Big)\, dx^i\, dx^k + \frac{\partial u_l}{\partial x^i}\frac{\partial u^l}{\partial x^k}\, dx^i\, dx^k$$

(assuming summation over repeated superscripts and subscripts $i, k, l$). The tensor

$$u_{ik} = \frac{1}{2}\Big(\frac{\partial u_i}{\partial x^k} + \frac{\partial u_k}{\partial x^i} + \frac{\partial u_l}{\partial x^i}\frac{\partial u^l}{\partial x^k}\Big)$$

is called the *strain tensor*.

In the linear approximation, which corresponds to small deformations, this tensor becomes

$$u_{ik} = \frac{1}{2}\Big(\frac{\partial u_i}{\partial x^k} + \frac{\partial u_k}{\partial x^i}\Big).$$

If the continuum is out of the equilibrium state, there arise forces that tend to return it to the equilibrium. It follows from physical considerations that these forces act on the given domain through its boundary surface, and the pressure acting on a small area orthogonal to a vector $n$ has the form $P(n)\, dS$, where $P = (P_{ik})$ is a linear operator,

$$P(n) = \sum_k P_{ik} n_k$$

(recall that all elements are written with subscripts because a single Euclidean coordinate system is fixed). The tensor $P_{ik}$ is called the *stress tensor*.

If the continuum satisfies Pascal's law, which says that pressure in the continuum is the same in all directions and depends only on the point, then the stress tensor is diagonal,

$$P_{ik} = p\delta_{ik},$$

where $p$ is the pressure at a point of the continuum.

According to Hooke's law, for small deformations of a continuum, the stress depends on deformation linearly. This is expressed by the following formula for the stress tensor:

$$P_{ik} = \sum_{l,m} T_{iklm} u_{lm}.$$

The tensor $T$ has 81 components, but for an isotropic continuum it must have the same form in all orthonormal coordinates. The hypothesis of isotropy

holds for fluids, but not always for solid bodies. Under the isotropy condition we can apply Theorem 7.8 for a description of the tensor $T$. We obtain

$$\sum_{l,m} T_{iklm}u_{lm} = \alpha u_{ik} + \beta u_{ki} + \gamma\delta_{ik}\operatorname{Tr} u,$$

where $\operatorname{Tr} u = \sum_l u_{ll}$. Since $u_{ik} = u_{ki}$, we may combine the first two terms to obtain the following form of Hooke's law:

$$P_{ik} = 2\mu u_{ik} + \lambda\delta_{ik}\operatorname{Tr} u.$$

The quantities $\lambda$ and $\mu$ involved in Hooke's law are called the *Lamé coefficients.*

We can conclude that in an isotropic continuum, a linear relationship between two symmetric tensors of rank two is specified by a tensor of rank 4, which is determined at each point by two constants (by the Lamé coefficients in the case of Hooke's law).

## 7.3. Exterior forms

**7.3.1. Symmetrization and alternation.** Two operations, symmetrization and alternation, are related to the operation of permuting indices.

Let $V^* \otimes V^*$ be the space of tensors of type $(0,2)$. These tensors are bilinear functions on $V$:

$$T(\xi,\eta) = T_{ij}\xi^i\eta^j.$$

Now, we have the group of permutations $S_2$ acting on the space of such tensors. This group consists of two elements, the identity permutation, preserving all tensors, and the permutation $\sigma = \left(\begin{smallmatrix}1&2\\2&1\end{smallmatrix}\right)$ transforming the tensor $T_{ij}$ into $\sigma(T)_{ij} = T_{ji}$. The space $V^*\otimes V^*$ decomposes then into the direct sum

$$V^*\otimes V^* = W^{\mathrm{sym}} \oplus W^{\mathrm{alt}},$$

and every permutation $\sigma \in S_2$ acts on these subspaces as multiplication by the following constants:

$$\sigma\big|_{W^{\mathrm{sym}}} = 1, \quad \sigma\big|_{W^{\mathrm{alt}}} = \operatorname{sgn}\sigma.$$

Indeed, each tensor can be represented in the form

$$T_{ij} = \frac{1}{2}(T_{ij}+T_{ji}) + \frac{1}{2}(T_{ij}-T_{ji}),$$

and the tensor

$$T_{ij}^{\mathrm{sym}} = \frac{1}{2}(T_{ij}+T_{ji})$$

is obtained from $T_{ij}$ by means of *symmetrization*, while the tensor

$$T_{ij}^{\mathrm{alt}} = \frac{1}{2}(T_{ij}-T_{ji}),$$

by means of *alternation*. It remains to note that the tensor $T^{\text{sym}}$ is symmetric, i.e., it does not change under permutations of indices (so that $T^{\text{sym}} \in W^{\text{sym}}$), and the tensor $T^{\text{alt}}$ is specified by a skew-symmetric matrix, $T^{\text{alt}}_{ij} = -T^{\text{alt}}_{ji}$, $T^{\text{alt}} \in W^{\text{alt}}$.

Let $e^1, \dots, e^n$ be a basis in $V^*$; then all the possible products $e^i \otimes e^j$, $i, j = 1, \dots, n$, form a basis in $V^* \otimes V^*$. Applying the operations of symmetrization and alternation to this basis, we conclude that the basis in the space of symmetric tensors of type $(0, 2)$ is

$$e^i \otimes e^j + e^j \otimes e^i, \quad i \leq j,$$

while the basis in the space of skew-symmetric tensors is

$$e^i \wedge e^j = e^i \otimes e^j - e^j \otimes e^i, \quad i < j.$$

Indeed, a symmetric tensor $T_{ij}$ is decomposed as

$$T_{ij} = \sum_i T_{ii} e^i \otimes e^i + \sum_{i<j} T_{ij}(e^i \otimes e^j + e^j \otimes e^i),$$

while a skew-symmetric tensor $T_{ij}$ as

$$T_{ij} = \sum_{i<j} T_{ij} e^i \wedge e^j.$$

These bases are written in the differential form as

$$dx^i\, dx^j = \frac{e^i \otimes e^j + e^j \otimes e^i}{2}, \quad dx^i \wedge dx^j = e^i \wedge e^j.$$

**7.3.2. Skew-symmetric tensors of type $(0, k)$.** The definition of skew-symmetry is extended to the case of arbitrary tensors of type $(0, k)$ as follows: a tensor $T_{i_1 \dots i_k}$ is said to be *skew-symmetric* if for each permutation $\sigma \in S_k$,

$$T_{i_{\sigma(1)}, \dots, i_{\sigma(k)}} = \operatorname{sgn}(\sigma) T_{i_1 \dots i_k}.$$

This means that the tensor changes its sign under odd permutations and remains unchanged under even permutations. This interpretation shows that the definition does not depend on the choice of coordinates.

**Theorem 7.10.** 1. *If the rank of a skew-symmetric tensor $T_{i_1 \dots i_k}$ is greater than the dimension of the vector space $V$, i.e., $k > \dim V$, then the tensor $T$ is identically equal to zero.*

2. *If $n$ is the dimension of the space, then there exists a single (up to proportionality) skew-symmetric tensor of type $(0, n)$.*

**Proof.** If at least two indices of a component of a skew-symmetric tensor $T_{i_1 \dots i_k}$ coincide, $i_p = i_q$, $p \neq q$, then this component is equal to zero. To show this, take the transposition $\sigma$ that interchanges $p$ and $q$. We know that $\operatorname{sgn}(\sigma) = -1$, hence $T_{i_1 \dots i_p \dots i_q \dots i_n} = -T_{i_1 \dots i_q \dots i_p \dots i_n} = 0$.

If the rank of $T$ is greater than the dimension of the space, then each component has at least two equal indices; hence this tensor is equal to zero. If the rank of $T$ is equal to the dimension of the space, then only for the components of the form $T_{\sigma(1)\dots\sigma(n)}$ are all indices different. But

$$T_{1\dots n} = \operatorname{sgn}(\sigma) T_{\sigma(1)\dots\sigma(n)},$$

hence the tensor is completely determined by the component $T_{1\dots n}$. □

We introduce the tensor $\varepsilon_{i_1\dots i_n}$ in $n$-dimensional space by the formula

$$\varepsilon_{i_1\dots i_n} = \operatorname{sgn}\begin{pmatrix} 1 & & n \\ i_1 & \dots & i_n \end{pmatrix}.$$

It is skew-symmetric and specifies a basis in the space of skew-symmetric tensors of maximal rank: any such tensor $T$ has the form

$$T_{i_1\dots i_n} = \varepsilon_{i_1\dots i_n} T_{1\dots n}.$$

For $n = 3$ we already introduced this tensor.

**Theorem 7.11.** *Let $T$ be a skew-symmetric tensor of rank $n$ in $n$-dimensional space with coordinates $x^1, \dots, x^n$, and let $z^i = z^i(x^1, \dots, x^n)$, $i = 1, \dots, n$, be a change of coordinates. Then in the new coordinate system the tensor $T$ is written as $\tilde{T}$, where*

$$\tilde{T}_{1\dots n} \det J = T_{1\dots n}$$

*and $J = \left(\frac{\partial z^i}{\partial x^j}\right)$ is the Jacobi matrix of the change of coordinates.*

**Proof.** By formula (7.7) of tensor transformation we have

$$T_{1\dots n} = \tilde{T}_{i_1\dots i_n} \frac{\partial z^{i_1}}{\partial x^1} \cdots \frac{\partial z^{i_n}}{\partial x^n} = \tilde{T}_{1\dots n}\left(\varepsilon_{i_1\dots i_n} \frac{\partial z^{i_1}}{\partial x^1} \cdots \frac{\partial z^{i_n}}{\partial x^n}\right) = \tilde{T}_{1\dots n} \det J.$$

Hence the theorem. □

A basis in the space of skew-symmetric tensors of type $(0, k)$ is given by the elements

$$e^{i_1} \wedge \dots \wedge e^{i_k} = dx^{i_1} \wedge \dots \wedge dx^{i_k}, \quad i_1 < \dots < i_k,$$

where

$$dx^{i_1} \wedge \dots \wedge dx^{i_k} = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma)\, dx^{\sigma(i_1)} \otimes \dots \otimes dx^{\sigma(i_k)}.$$

Obviously, each expression $dx^{i_1} \wedge \dots \wedge dx^{i_k}$ is skew-symmetric relative to permutations of indices,

$$dx^{\sigma(i_1)} \wedge \dots \wedge dx^{\sigma(i_k)} = \operatorname{sgn}(\sigma)\, dx^{i_1} \wedge \dots \wedge dx^{i_k},$$

and skew-symmetric tensors are expanded in this basis as follows:

$$T_{i_1\dots i_k}\, dx^{i_1} \otimes \dots \otimes dx^{i_k} = \sum_{i_1 < \dots < i_k} T_{i_1\dots i_k}\, dx^{i_1} \wedge \dots \wedge dx^{i_k}.$$

The expressions

$$\omega = \sum_{i_1 < \cdots < i_k} T_{i_1 \ldots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

are also called *exterior* or *differential forms*. The rank of the skew-symmetric tensor is referred to as the *degree* of the differential form: $\deg \omega = k$.

There are also objects which behave as skew-symmetric tensors only under coordinate changes with positive Jacobian, $\det J > 0$. We will demonstrate here an important example.

Let $g_{ij}\, dx^i\, dx^j$ be a Riemannian metric in the space. Consider the expression

$$\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n,$$

where $g = \det(g_{ij})$. It is called the *volume element* determined by the metric $g_{ij}\, dx^i\, dx^j$.

**Theorem 7.12.** *The volume element $\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n$ changes as a tensor under coordinate changes such that $\det J = \det\left(\frac{\partial z^i}{\partial x^j}\right) > 0$, i.e., under coordinate changes with positive Jacobian.*

**Proof.** On changing to new coordinates $(z)$ the metric becomes $\tilde{g}_{ij}\, dz^i\, dz^j$, where

$$\tilde{g}_{ij} = \frac{\partial x^k}{\partial z^i} \frac{\partial x^l}{\partial z^j} g_{kl}.$$

In matrix terms this means that the matrix $\tilde{G} = (\tilde{g}_{kl})$ has the form

$$\tilde{G} = A^\top G A,$$

where $A = J^{-1} = \left(\frac{\partial x^j}{\partial z^l}\right)$ and $A^\top$ is its transpose. This implies that $\det \tilde{G} = \det G\, (\det A)^2$, whence

$$|\det J| \sqrt{\tilde{g}} = \sqrt{g}.$$

□

**7.3.3. Exterior algebra. Symmetric algebra.** It is easily seen that the tensor product of skew-symmetric tensors is not skew-symmetric. For such tensors there is another operation resulting in a skew-symmetric tensor.

Let $R_{i_1 \ldots i_p}$ and $S_{j_1 \ldots j_q}$ be skew-symmetric tensors. They are associated with differential forms

$$\omega_1 = \sum_{i_1 < \cdots < i_p} R_{i_1 \ldots i_p}\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}, \quad \omega_2 = \sum_{j_1 < \cdots < j_q} S_{j_1 \ldots j_q}\, dx^{j_1} \wedge \cdots \wedge dx^{j_q}.$$

Define their *exterior product* as the form

$$\omega_1 \wedge \omega_2 = \sum_{k_1 < \cdots < k_{p+q}} T_{k_1 \ldots k_{p+q}}\, dx^{k_1} \wedge \cdots \wedge dx^{k_{p+q}},$$

where the tensor $T$ is defined by

$$T_{k_1\dots k_{p+q}} = \frac{1}{p!q!}\sum_{\sigma\in S_{p+q}} \operatorname{sgn}(\sigma) R_{\sigma(k_1)\dots\sigma(k_p)} S_{\sigma(k_{p+1})\dots\sigma(k_{p+q})}. \tag{7.11}$$

Actually, we have already introduced the exterior product of basis forms $dx^i$ when we defined the forms $dx^{i_1}\wedge\dots\wedge dx^{i_k}$.

**Theorem 7.13.** *The exterior product of differential forms is associative,*

$$(\omega_1\wedge\omega_2)\wedge\omega_3 = \omega_1\wedge(\omega_2\wedge\omega_3),$$

*bilinear in both arguments, and skew-commutative, i.e.,*

$$\omega_1\wedge\omega_2 = (-1)^{pq}\omega_2\wedge\omega_1, \tag{7.12}$$

*where* $p = \deg\omega_1$, $q = \deg\omega_2$.

**Proof.** Bilinearity and associativity follow in an obvious way from (7.11). To prove (7.12), it suffices to note that the sign of the permutation

$$\begin{pmatrix} 1 & & q & q+1 & & p+q \\ p+1 & & p+q & 1 & & p \end{pmatrix}$$

is $(-1)^{pq}$. □

**Corollary 7.1.** 1. *The skew-symmetric tensors in a domain $U\subset\mathbb{R}^n$ or on a regular surface $M$ form a graded-commutative algebra relative to the operation of exterior product.*

2. *Let $V$ be the tangent space at a point. Then the exterior product determines in the space of forms at this point the structure of exterior algebra (or Grassmann's algebra, or complete exterior power of the space $V^*$).*

Since every vector space $L$ can be represented as the space of covectors at a point, one can formally introduce the exterior product on this space. The algebra generated by vectors of the space $L$ (for example, $L = V^*$) relative to the exterior product is denoted by $\Lambda^*L$. It is graded by the degrees of forms:

$$\Lambda^*L = \bigoplus_{k=0}^{n}\Lambda^k L, \quad n = \dim L,$$

where $\Lambda^k L$ is the subspace consisting of all forms of degree $k$.

Let $L$ be a Lie algebra. We introduce the bilinear operation $\{\ ,\}$ on its exterior power $\Lambda^*L$ by the following rule:

1) if $v, w\in L$, then $\{v, w\} = [v, w]$ (this operation coincides with the commutator in the Lie algebra);

2) if $a \in L$, $v \in \Lambda^k L$, $w \in \Lambda^m L$, then

$$\{a \wedge v, w\} = a \wedge \{v, w\} + (-1)^k v \wedge \{a, w\},$$
$$\{v, w \wedge a\} = \{v, w\} \wedge a + (-1)^m \{v, a\} \wedge w.$$

These conditions uniquely determine the operation which is specified on homogeneous vectors in $\Lambda^* L$ by the following formula:

$$\{v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_m\}$$
$$= \sum_{i,j} (-1)^{i+j} [v_i, w_j] \wedge v_1 \cdots \wedge \widehat{v_i} \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge \widehat{w_j} \wedge \cdots \wedge w_m,$$

where the "hat", as usual, means that this symbol is skipped. We see that

$$\{v, w\} \in \Lambda^{k+m-1} L \quad \text{for } v \in \Lambda^k L,\ w \in \Lambda^m L.$$

Decompose the algebra $\Lambda^* L$ into the direct sum $L_0 \oplus L_1$, where $L_0$ is the linear span of all the vectors in $\Lambda^{2k+1} L$, and $L_1$ is the linear span of all the vectors in $\Lambda^{2k} L$, $k = 0, 1, \ldots$. Set $\deg u = \alpha$ for $u \in L_\alpha$. We leave it as an exercise to prove that the operation thus obtained specifies the structure of *Lie superalgebra* on $\Lambda^* L$:

$$\{u, v\} \in L_{(\alpha+\beta) \bmod 2},$$
$$\{u, v\} = (-1)^{\alpha\beta+1} \{v, u\},$$
$$(-1)^{\alpha\gamma} \{u, \{v, w\}\} + (-1)^{\beta\alpha} \{v, \{w, u\}\} + (-1)^{\gamma\beta} \{w, \{u, v\}\} = 0,$$

where $\deg u = \alpha$, $\deg v = \beta$, $\deg w = \gamma$.

For polyvector fields (skew-symmetric tensors with superscripts; see Section 7.5.1) this operation is called the *Schouten bracket*; for vector fields it becomes the usual commutator.

## 7.4. Tensors in the space with scalar product

**7.4.1. Raising and lowering indices.** Let $g_{ij}$ be a tensor of type $(0, 2)$ that defines a nondegenerate scalar product of vectors,

$$\langle \xi, \eta \rangle = g_{ij} \xi^i \eta^j.$$

If the tensor $g_{ij}$ is symmetric, then it specifies a Riemannian or a pseudo-Riemannian metric, whereas a skew-symmetric tensor specifies a symplectic scalar product. Since the scalar product is nondegenerate, the matrix $(g_{ij})$ has a nonzero determinant. Therefore, there exists the inverse tensor $g^{kl}$ of type $(2, 0)$ defined by the condition

$$g_{ik} g^{kj} = g^{jk} g_{ki} = \delta_i^j.$$

It defines the scalar product of covectors:

$$\langle \xi, \eta \rangle = g^{ij} \xi_i \eta_j.$$

Having a scalar product $g_{ij}$, we can define the operation of lowering the index. It associates with each tensor $T^{i_1 \dots i_p}_{j_1 \dots j_q}$ of type $(p, q)$ the tensor $\hat{T}^{i_2 \dots i_p}_{i_1 j_1 \dots j_q}$ of type $(p-1, q+1)$ by the following rule:

$$\hat{T}^{i_2 \dots i_p}_{i_1 j_1 \dots j_q} = g_{i_1 k} T^{k i_2 \dots i_p}_{j_1 \dots j_q}. \tag{7.13}$$

**Lemma 7.3.** *The operation* (7.13) *results again in a tensor.*

**Proof.** This operation can be represented as a composition of two operations, which have already been shown to transform tensors into tensors. Namely, consider first the tensor product $S = g \otimes T$:

$$S^{l i_2 \dots i_p}_{i_1 l' j_1 \dots j_q} = g_{i_1 l'} T^{l i_2 \dots i_p}_{j_1 \dots j_q},$$

and then take the contraction of this tensor in the superscript $l$ and the subscript $l'$. As a result, we obtain the tensor $\hat{T}$. ☐

The passage from the tensor $T^{i_1 \dots i_p}_{j_1 \dots j_q}$ to $T^{i_2 \dots i_p}_{i_1 j_1 \dots j_q}$ is called *lowering the index* $i_1$ by means of the scalar product $g_{ij}$. Usually, the case of a metric $g_{ij}$ is considered, but in the framework of the Hamiltonian formalism it is expedient to use this operation in the case of skew-symmetric scalar products.

EXAMPLE. If $\xi^i$ is a vector, then lowering the index yields a covector

$$\xi_i = g_{ij} \xi^j.$$

Therefore, lowering an index specifies a linear mapping of the space of vectors into the space of covectors. With a vector $\xi$ it associates the covector $\hat{\xi}$ such that for any vector $\eta$ the scalar product $\langle \xi, \eta \rangle$ equals the value of the functional $\hat{\xi}$ on the vector $\eta$:

$$\hat{\xi}(\eta) = \langle \xi, \eta \rangle.$$

In a similar way the inverse tensor $g^{ij}$ specifies the operation of *raising the index* by the rule

$$\hat{T}^{j_1 i_1 \dots i_p}_{j_2 \dots j_q} = g^{j_1 k} T^{i_1 \dots i_p}_{k j_2 \dots j_q}.$$

**Lemma 7.4.** *If we lower an index and then raise it, we obtain the initial tensor,* $\hat{\hat{T}} = T$.

**Proof.** This follows from the formula

$$g^{i_1 l} \hat{T}^{i_2 \dots i_p}_{l j_1 \dots j_q} = g^{i_1 l} g_{lk} T^{k i_2 \dots i_p}_{j_1 \dots j_q} = \delta^{i_1}_k T^{k i_2 \dots i_p}_{j_1 \dots j_q} = T^{i_1 \dots i_p}_{j_1 \dots j_q}.$$

☐

The matrix $g^{ij}$ specifies a scalar product on the space of covectors. The following lemma shows that lowering an index of a vector preserves scalar products.

**Lemma 7.5.** *Let $\xi = (\xi^i)$ and $\eta = (\eta^i)$ be vectors, and let $\hat{\xi} = (\hat{\xi}_i) = (g_{ij}\xi^j)$ and $\hat{\eta} = (\hat{\eta}_i) = (g_{ij}\eta^j)$ be the covectors obtained from $\xi$ and $\eta$ by lowering the index. Then their scalar products specified by the tensors $g_{ij}$ and $g^{ij}$ satisfy the relation*

$$\langle \hat{\xi}, \hat{\eta} \rangle = \langle \eta, \xi \rangle.$$

*In particular, they coincide for a symmetric scalar product.*

**Proof.** The scalar product of covectors can be written as

$$\langle \hat{\xi}, \hat{\eta} \rangle = g^{ij}\hat{\xi}_i\hat{\eta}_j = g^{ij}g_{ik}\xi^k g_{jl}\eta^l = \delta^i_l g_{ik}\xi^k\eta^l = \eta^i g_{ik}\xi^k = \langle \eta, \xi \rangle.$$

Hence the lemma. □

Consider the case where $g_{ij}$ is a metric in Euclidean space. In Euclidean coordinates it has the form $g_{ij} = \delta_{ij}$. Therefore, lowering (or raising) an index does not change the components of any tensor (relative to Euclidean coordinates),

$$\hat{T}^{i_2 \ldots i_p}_{i_1 j_1 \ldots j_q} = \delta_{i_1 i} T^{i i_2 \ldots i_p}_{j_1 \ldots j_q} = T^{i_1 i_2 \ldots i_p}_{j_1 \ldots j_q}.$$

Hence there is no difference between superscripts and subscripts in Euclidean coordinates. For this reason all tensors are usually written with subscripts. This form is invariant with respect to transformations that preserve the Euclidean metric (orthogonal transformations). For example, the gradient behaves like a vector under such transformations.

**7.4.2. Eigenvalues of scalar products.** Suppose in a vector space there is a scalar product specified by a tensor $T_{ij}$ and a metric specified by a tensor $g_{ij}$. The *eigenvalues* of the scalar product $T_{ij}$ in the metric $g_{ij}$ are defined as the solutions to the equation

$$P(\lambda) = \det(T_{ij} - \lambda g_{ij}) = 0.$$

This definition of eigenvalues is invariant with respect to the choice of coordinates.

EXAMPLE. Let $b_{ij}$ be the second fundamental form of a surface and $g_{ij}$ the first fundamental form of the same surface. Then the principal curvatures $k_1$ and $k_2$ are the eigenvalues of the second fundamental form in the metric $g_{ij}$.

Without a metric, it makes no sense to speak about symmetry or skew-symmetry of linear operators (tensors of type $(1,1)$). Given a metric $g_{ij}$, a

linear operator $T^i_j$ in the space with this metric is said to be *symmetric* (or *skew-symmetric*) if the bilinear form

$$T_{ij} = g_{ik}T^k_j$$

is symmetric ($T_{ij} = T_{ji}$) or skew-symmetric ($T_{ji} = -T_{ij}$).

The definition of the quadratic form $T_{ij}$ implies the following result.

**Lemma 7.6.** *A linear operator $T = (T^i_j)$ in the space with metric $g_{ij}$ (possibly, pseudo-Euclidean) is symmetric or skew-symmetric if and only if for all vectors $\xi$ and $\eta$ the following identities hold:*

$$\langle T\xi, \eta\rangle = \langle \xi, T\eta\rangle \quad (\textit{symmetry})$$

*or*

$$\langle T\xi, \eta\rangle = -\langle \xi, T\eta\rangle \quad (\textit{skew-symmetry}).$$

Suppose the following equalities hold:

$$T^i_j\xi^j = \lambda\xi^i, \quad i = 1, \dots, n,$$

i.e., $\xi$ is an eigenvector of the operator $T^i_j$ with eigenvalue $\lambda$. Then

$$T_{ik}\xi^k = g_{ij}T^j_k\xi^k = \lambda g_{ik}\xi^k, \quad i = 1, \dots, n,$$

i.e., we have

$$(T_{ik} - \lambda g_{ik})\xi^k = 0, \quad i = 1, \dots, n.$$

The trace

$$\operatorname{Tr} T^i_j = T^i_i = g^{ik}T_{ki}$$

and the determinant $\det T^i_j$ of the linear operator $T^i_j = g^{ik}T_{kj}$ are invariant under the coordinate changes preserving the metric, but, in general, they depend on the metric.

EXAMPLE. For a regular surface in $\mathbb{R}^3$, the Gaussian curvature $K$ and the mean curvature $H$ are defined by the formulas

$$K = \frac{\det(b_{ij})}{\det(g_{ij})}, \quad H = \frac{1}{2}\, b^i_i = \frac{1}{2}\, g^{ij}b_{ij},$$

where $g_{ij}$ and $b_{ij}$ are the first and the second fundamental forms of the surface. Hence

$$K = (\det(g_{ij}))^{-1}\det(b_{ij}) = \det(g^{ik}b_{kj}) = \det(b^i_j).$$

Thus we see that $K$ and $2H$, being the determinant and the trace of the operator $b^i_j$, are metric invariants of the second fundamental form.

**7.4.3. Hodge duality operator.** In the $n$-dimensional space with metric, there is a certain duality operator (the Hodge duality operator, also called the Hodge star-operator) that assigns to a skew-symmetric tensor $T$ of type $(0,k)$ a skew-symmetric tensor $*T$ of type $(0,n-k)$ by the formula

$$(*T)_{i_{k+1}\dots i_n} = \frac{1}{k!}\sqrt{|g|}\varepsilon_{i_1\dots i_n}T^{i_1\dots i_k},$$

where

$$T^{i_1\dots i_k} = g^{i_1 j_1}\cdots g^{i_k j_k}T_{j_1\dots j_k}.$$

Since $\sqrt{|g|}\varepsilon_{i_1\dots i_n}$ is a tensor relative to coordinate changes with positive Jacobian (see Section 7.3.2), $*T$ is also a tensor relative to such coordinate changes. Obviously, $*T$ is skew-symmetric.

Note that the duality operator is linear,

$$*(S+T) = *S + *T, \quad *(fT) = f * T,$$

where $f$ is a function on a manifold.

EXAMPLE. Let $g_{ij}$ be a Riemannian (positive definite) metric that has the Euclidean form $g_{ij} = \delta_{ij}$ at any point $x_0$, and let the basis of tangent vectors $(\frac{\partial}{\partial x^1},\dots,\frac{\partial}{\partial x^n})$ is positively oriented. Then, at this point,

$$*(f(x)\,dx^1\wedge\cdots\wedge dx^k) = f(x)\,dx^{k+1}\wedge\cdots\wedge dx^n.$$

In particular, for three-dimensional Euclidean space with coordinates $x, y$, and $z$, we have

$$*dx = dy\wedge dz, \quad *dy = -dx\wedge dz, \quad *dz = dx\wedge dy,$$

and by linearity

$$*\omega = P\,dy\wedge dz + Q\,dz\wedge dx + R\,dx\wedge dy$$

for an arbitrary 1-form $\omega = P\,dx + Q\,dy + R\,dz$. Note also that

$$*(*\omega) = \omega,$$

and if $f$ is a scalar (function), then $*f = f\,dx\wedge dy\wedge dz$ is a form of rank 3.

The proof of the following formula is left as an exercise for the reader:

$$*(*T) = (-1)^{k(n-k)}\operatorname{sgn}(g)T. \tag{7.14}$$

**7.4.4. Fermions and bosons. Spaces of symmetric and skew-symmetric tensors as Fock spaces.** Consider a space with metric $g_{ij} = \text{const}$. In this section we assume that this metric is Euclidean, i.e., symmetric and positive. For convenience, we also assume that the metric is given in an orthonormal basis $e_1, \ldots, e_n \in \mathbb{R}^n$, where $g_{ij} = \langle e_i, e_j \rangle = \delta_{ij}$.

According to physics conceptions of the 20th century, quantum particles are associated with operators, and the particles fall into two types: fermions and bosons. We will describe here a very simple algebra lying behind these concepts, which is sometimes also very useful within geometry itself.

We introduce the associative algebra generated by elements

$$\begin{gathered} a_1, \ldots, a_k, \quad b_1, \ldots, b_l, \\ a_1^+, \ldots, a_k^+, \quad b_1^+, \ldots, b_l^+ \end{gathered}$$

and relations

$$\begin{gathered} a_i a_j + a_j a_i = 0, \quad a_i^+ a_j^+ + a_j^+ a_i^+ = 0, \\ a_i a_j^+ + a_j^+ a_i = g_{ij} = \delta_{ij}, \\ [a_i, b_j] = [a_i, b_j^+] = [b_i, b_j] = [b_i^+, b_j^+] = 0, \\ [b_i, b_j^+] = g_{ij} = \delta_{ij}, \end{gathered}$$

where $[a, b] = ab - ba$.

The most fundamental representation (due to Dirac), where this algebra is realized as an operator algebra, is constructed as follows. We introduce the so-called *vacuum* (or generating) vector $|0\rangle = \eta_0$ and require the following relations to hold:

$$a_i \eta_0 = b_j \eta_0 = 0, \quad i = 1, \ldots, k, \quad j = 1, \ldots, l.$$

We will call the operators $a_i$ and $b_j$ the *annihilation operators* of particles, specifically, of fermions for $a_i$ and bosons for $b_j$. Now we assume that no more relations are given: all relations must follow from the commutation relations in the algebra and the properties of the vacuum vector $\eta_0 = |0\rangle$. In other words, this means that all the basis vectors must be obtained from the vacuum $\eta_0 = |0\rangle$ by successive application of *creation operators* of fermions $a_j^+$ and bosons $b_j^+$:

$$a_{i_1}^+ \cdots a_{i_q}^+ b_{j_1}^+ \cdots b_{j_p}^+ \eta_0.$$

The space generated by the action of the algebra of operators $a_i$, $a_i^+$, $b_j$, $b_j^+$ on the vector $\eta_0$ will be called the *Fock space* $F_{k,l}$.

We introduce a scalar product in the Fock space subject to the following requirements:

1) $|\eta_0| = 1$.

2) The vectors

$$(a_1^+)^{\alpha_1} \cdots (a_k^+)^{\alpha_k} (b_1^+)^{m_1} \cdots (b_l^+)^{m_l} \eta_0$$

(see above) form an orthogonal basis in the Fock space for all tuples $(\alpha, m)$ such that $\alpha_j = 0, 1$, $m_j = 0, 1, 2, 3, \ldots$.

3) The annihilation operators $b_j$, $a_i$ are dual to $b_j^+$, $a_i^+$.

These requirements completely determine the scalar product. We will construct the Fock space successively, based on the relations $a_i|0\rangle = b_j|0\rangle = 0$. Using the commutation relations we obtain the following assertion.

**Lemma 7.7.** *Any vector obtained from the vacuum $\eta_0$ by application of operators $a_i$, $a_i^+$, $b_j$, $b_j^+$ can be reduced to a linear combination of vectors that involve no symbols $a_i$ and $b_j$.*

**Proof.** Any operator $a_i$ commutes with $b_j$ and $b_j^+$ for all $j$ and anticommutes with all $a_q$ (i.e., $a_i a_q = -a_q a_i$) and all $a_p^+$ for $p \neq i$ (i.e., $a_i a_p^+ = -a_p^+ a_i$). For $p = i$ we have $a_p a_p^+ = 1 - a_p^+ a_p$. Moreover, from the same relations we obtain $a_i^2 = (a_i^+)^2 = 0$. Consider any vector of the form $\eta = (\ldots) a_i (\ldots)|0\rangle$, where the parentheses embrace words consisting of the letters $a_q$, $a_p^+$, $b_j$, $b_j^+$. Applying the above commutation relations, we move $a_i$ to the right, "closer to the beginning" $|0\rangle$. Performing this operation repeatedly, $a_i$ either arrives at the beginning, in which case we use the relation $(\ldots)a_i|0\rangle = 0$, or $a_i$ meets the element $a_i^+$. In the latter case we use the relation $a_i a_i^+ = -a_i^+ a_i + 1$, which represents our element as a linear combination of two words in which $a_i$ either comes closer to the beginning (left) or disappears (right). Thus we finally arrive at a linear combination of words which do not contain $a_i$. In the same way we obtain a linear combination of words containing no symbols $b_j$. $\square$

Next, all the words can be reduced to the form where the letters are written in the natural order:

$$(a_1^+)^{\alpha_1} \cdots (a_k^+)^{\alpha_k} (b_1^+)^{m_1} \cdots (b_l^+)^{m_l} |0\rangle.$$

Obviously, $\alpha_i = 0, 1$, since $(a_i^+)^2 = 0$. The $m_j$ may be any numbers, $m_j \geq 0$. This set of vectors is linearly independent.

Now we define the scalar product in the Fock space $F_{k,l}$.

**Lemma 7.8.** *Suppose that in the Fock space there is a scalar product such that $\langle \eta_0, \eta_0 \rangle = 1$ and the pairs of operators $(a_i, a_i^+)$, $(b_i, b_i^+)$ are formally conjugate. Then the basis constructed above is orthogonal and*

$$\langle \eta, \eta \rangle = m_1! \cdots m_l!,$$

*where* $\eta = a_1^{+\alpha_1} \cdots a_k^{+\alpha_k} b_1^{+m_1} \cdots b_l^{+m_l} \eta_0$.

**Proof.** We proceed to construction of the Fock space. For the action of $a_i^+$ we have

$$\langle a_i^+\eta_0, \eta_0\rangle = \langle \eta_0, a_i\eta_0\rangle = 0,$$
$$\langle a_i^+\eta_0, a_i^+\eta_0\rangle = \langle \eta_0, a_i a_i^+\eta_0\rangle = \langle \eta_0, \eta_0\rangle - \langle \eta_0, a_i^+ a_i\eta_0\rangle = \langle \eta_0, \eta_0\rangle = 1.$$

In a similar way, for the action of operators $b_i^+$ we have

$$\langle b_i^+\eta_0, \eta_0\rangle = \langle \eta_0, b_i\eta_0\rangle = 0,$$
$$\langle b_i^+\eta_0, b_i^+\eta_0\rangle = \langle \eta_0, b_i b_i^+\eta_0\rangle = \langle \eta_0, \eta_0\rangle + \langle \eta_0, b_i^+ b_i\eta_0\rangle = \langle \eta_0, \eta_0\rangle = 1,$$
$$\langle (b_i^+)^2\eta_0, (b_i^+)^2\eta_0\rangle = \langle b_i^+\eta_0, b_i(b_i^+)^2\eta_0\rangle = \langle b_i^+\eta_0, b_i^+\eta_0\rangle + \langle b_i^+\eta_0, b_i^+\eta_0\rangle = 2,$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\langle (b_i^+)^m\eta_0, (b_i^+)^m\eta_0\rangle = m!.$$

If we continue by induction to calculate the scalar product, we obtain the formulas stated above. □

Now we have the following theorem.

**Theorem 7.14.** *The Fock space $F_{k,l}$ is well defined as a direct sum of finite-dimensional Euclidean spaces with basis consisting of vectors of the form $\eta = a_1^{+\alpha_1}\cdots a_k^{+\alpha_k} b_1^{+m_1}\cdots b_l^{+m_l}\eta_0$, where $\alpha_j = 0, 1$ and $m_j = 0, 1, 2, \dots$ are nonnegative integers.*

This basis is orthogonal, and the squared norm of a basis vector as stated above equals $m_1!\cdots m_l!$. The annihilation operators $a_i$ are dual to $a_i^+$, the operators $b_i$ are dual to $b_i^+$, and $a_i\eta_0 = b_j\eta_0 = 0$ for all $i, j$. We introduce the following grading on the vectors $\eta$ of the form stated above:

$$\eta \to \left(\sum \alpha_j, \sum m_j\right) = (\alpha, m).$$

**Lemma 7.9.** *The set of vectors with a given pair of numbers $(\alpha, m)$ generates a Euclidean vector space isomorphic to the tensor product*

$$\Lambda^\alpha \mathbb{R}^k \otimes S^m \mathbb{R}^l = F_{k,l}^{(\alpha,m)},$$

*where $\Lambda^\alpha\mathbb{R}^k$ is the exterior power of the space $\mathbb{R}^k$ (i.e., the set of all skew-symmetric tensors of rank $\alpha$ in $\mathbb{R}^k$), and $S^m\mathbb{R}^l$ is the symmetric power of the space $\mathbb{R}^l$ (i.e., the set of all symmetric tensors of rank $m$ on $\mathbb{R}^l$). Moreover, we have*

$$F_{k,l} = \bigoplus_{(\alpha,m)} F_{k,l}^{(\alpha,m)},$$

*where $\alpha = 0, 1, \dots, k$, and $m = 0, 1, 2, \dots$ .*

The *proof* of the lemma follows from the commutation relations in an obvious way.

Recall that all the symbols $a_i^+$ anticommute with each other, while the symbols $b_j^+$ commute.

EXAMPLE. Let $\eta_0 = ce^{-x^2/2}$, $x \in \mathbb{R}$, where the constant $c$ is chosen so that

$$\langle \eta_0, \eta_0 \rangle = \int_{-\infty}^{\infty} \eta_0^2 \, dx = \int_{-\infty}^{\infty} c^2 e^{-x^2} \, dx = 1.$$

Then $c^2 = \frac{1}{2\pi}$.

Next, let $-\sqrt{2}b_1 = d/dx + x$, $\sqrt{2}b_1^+ = d/dx - x$. Then

$$b_1\eta_0 = 0, \quad b_1 b_1^+ = 1 + b_1^+ b_1.$$

The basis of the space $F_{0,1}$ consists of the vectors

$$(b^+)^n \eta_0 = cP_n(x)e^{-x^2/2},$$

where the $P_n(x)$ are the Hermite polynomials. This is the well-known realization of the simplest boson Fock space in analysis.

The fermion Fock space is naturally realized in geometry as the space of exterior forms at a given point of a manifold.

Now we give the simplest examples of important operators acting in the Fock spaces of bosons and fermions.

1) The operators of the number of particles of a given kind $i$:

$$n_i^F = a_i^+ a_i \quad \text{(fermions)}$$

or

$$n_j^B = b_j^+ b_j \quad \text{(bosons)}.$$

2) The operator of the total number of particles, $\hat{n} = \sum_i n_i^F + \sum_j n_j^B$.

We have the following simple lemma.

**Lemma 7.10.** *The eigenvalues of the operators $n_i^F$ and $n_j^B$ acting on the Fock spaces are nonnegative integers, which for the operators $n_i^F$ are equal to 0 or 1. By definition, these numbers are equal to the number of fermions (respectively, bosons) in the given state.*

*The eigenvectors are exactly the iterations of the creation operators applied to the vacuum vector $|0\rangle = \eta_0$.*

*All the operators $n_i^F$ and $n_j^B$ commute with each other. Their sum $\hat{n} = \sum_i n_i^F + \sum_j n_j^B$ has the same eigenvectors and is called the "number-of-particles operator".*

**Proof.** The proof of this lemma is almost obvious. For example, applying the operator $n_i^F = a_i^+ a_i$ to any vector of the form $\eta = b_{j_1}^+ \cdots b_{j_k}^+ a_{i_1}^+ \cdots a_{i_l}^+ |0\rangle$, we obtain

$$a_i^+ a_i \eta = n_i^F \eta = 0$$

if $a_{i_q} \neq a_i$ for all $i_q$, since the operator $a_i$ commutes with all the operators $b_j^+$, anticommutes with all the operators $a_{i_q}^+$ for $i_q \neq i$ and, by definition, $a_i|0\rangle = 0$. Next, if one of the numbers coincides with $i$ (i.e., $i_q = i$), then we have $a_i a_i^+ = -a_i^+ a_i + 1$. Hence in this case $n_i^F \eta = a_i^+ a_i \eta = \eta$. This proves the lemma for fermions.

The situation for bosons when $j = j_q$ is similar. The vector $\eta$ has the form $(b_j^+)^{m_j}(\dots)|0\rangle$, where $m_j \geq 0$ and the part in the parentheses does not contain $b_j^+$. Applying the operator $n_j^B = b_j^+ b_j$ to the vector $\eta = (b_j^+)^{m_j}(\dots)|0\rangle$, we obtain

$$(b_j^+ b_j)(b_j^+)^{m_j}(\dots)|0\rangle = m_j (b_j^+)^{m_j}(\dots)|0\rangle = m_j \eta,$$

since $b_j|0\rangle = 0$ and $b_j$ commutes with all operators $b_k^+$ for $k \neq j$. The proof is completed. □

Thus the Fock space is a graded vector space, where the corresponding subspace of grade $n$ is exactly the eigenspace of the number-of-particles operator $\hat{n}$ with eigenvalue $n$, where $n = 0, 1, 2, \dots$ .

The simplest class of important operators acting in the Fock space is the class of quadratic forms, which are operators with quadratic expressions in terms of annihilation and creation operators $a_i^+$, $a_i$, $b_j^+$, $b_j$. For example, the number-of-particles operator $\hat{n}$ introduced above is an operator of this kind.

We define a *real fermion* or *boson quadratic form* as a selfadjoint operator that is a quadratic form of the operators of annihilation and creation of fermions or bosons, respectively, with real coefficients.

By definition such an operator has the form (with real coefficients)

$$A = \sum_{i,j} (u_{ij} a_i^+ a_j^+ + 2 v_{ij} a_i^+ a_j + w_{ij} a_i a_j) + C$$

(for fermions), where $u_{ij} = -u_{ji}$, $v_{ij} = v_{ji}$, and $w_{ij} = -u_{ij}$, since $A^+ = A$ (the operator $A$ is selfadjoint), or

$$B = \sum_{i,j} (u_{ij} b_i^+ b_j^+ + 2 v_{ij} b_i^+ b_j + w_{ij} b_i b_j) + C$$

(for bosons), where $u_{ij} = u_{ji}$, $v_{ij} = v_{ji}$, and $w_{ij} = u_{ij}$, since $B^+ = B$.

To satisfy the requirements of quantum theory and geometry, we must be able to reduce the fermion and boson quadratic forms to a diagonal form by means of the *Bogolyubov transformations*:

$$a_i = P_{ij}a_j'^{+} + Q_{ij}a_j', \quad b_i = P_{ij}b_j'^{+} + Q_{ij}b_j',$$
$$a_i^{+} = Q_{ij}a_j'^{+} + P_{ij}a_j', \quad b_i^{+} = Q_{ij}b_j'^{+} + P_{ij}b_j'$$

(where summation over $j$ is implied). We will require that the coefficients of the Bogolyubov transformation be real and that the new operators $a_i'^{+}, a_i'$ satisfy the same commutation relations as the ones we had before for creation and annihilation operators, with operators $a_i'^{+}$ and $a_i'$ being mutually conjugate:

$$[a_j'^{+}, a_k']_+ = \delta_{jk}, \quad [a_j'^{+}, a_k'^{+}]_+ = [a_j', a_k']_+ = 0, \quad \text{where } [a,b]_+ = ab + ba.$$

The only difference in commutation relations for bosons is the use of $[a,b]_- = ab - ba$ instead of $[\cdot,\cdot]_+$.

The following lemma is obvious.

**Lemma 7.11.** *The real fermion and boson quadratic forms can be written in the following "standard" real form:*

$$A = R_{ij}(a_i^{+} + a_i)(a_j^{+} - a_j) + C - \operatorname{Tr} R,$$
$$B = S^{ij}(b_i^{+} + b_i)(b_j^{+} + b_j) + T_{ij}(b_i^{+} - b_i)(b_j^{+} - b_j) + C - \operatorname{Tr} S + \operatorname{Tr} T,$$

*where*

$$R_{ij} = u_{ij} + v_{ij} \quad (\textit{fermions}),$$
$$S^{ij} + T_{ij} = u_{ij}, \quad S^{ij} - T_{ij} = v_{ij} \quad (\textit{bosons}).$$

In the case of bosons, this form has a natural meaning in terms of the operator realization of the Fock space. As before, we set

$$\eta_0 = \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{2}\right),$$
$$b_i = \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial x^i} + x^i\right), \quad b_i^{+} = \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial x^i} - x^i\right), \quad b_i\eta_0 = 0.$$

Then the operators $b_i^{+} + b_i$ and $b_i^{+} - b_i$ are exactly the operators

$$b_i^{+} + b_i = \sqrt{2}\,\frac{\partial}{\partial x^i}, \quad b_i^{+} - b^i = \sqrt{2}x^i.$$

Therefore, $S^{ij}(b_i^{+} + b_i)(b_j^{+} + b_j) = K$ is the kinetic (differential) part of the Schrödinger operator $B$, whereas $T_{ij}(b_i^{+} - b_i)(b_j^{+} - b_j)$ is a quadratic potential of the form

$$U(x) = 2\sum_{i,j} T_{ij}x^i x^j, \quad K = 2\sum_{i,j} S^{ij}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}.$$

Thus we obtained the standard *quantum oscillator* $B$. This operator is diagonalized by a single linear transformation of the $x$-space, which determines dual transformations in the spaces $(\frac{\partial}{\partial x})$ and $(x)$. We must reduce to the diagonal form a pair of quadratic forms one of which is a form on the space $\mathbb{R}^n$, and the other, on the dual space. As we know, two forms may be diagonalized by a linear transformation if one of these forms is positive, and the more so if both forms are positive. As a result, we may assume that

$$S^{ij} \to \delta_{ij} = S'^{ij}, \quad T_{ij} \to t_i^2 \delta_{ij} = T'_{ij}.$$

Let us see what this problem looks like in terms of the Bogolyubov transformations.

**Lemma 7.12.** *Under the real Bogolyubov transformation*

$$b_i = P_{ij} b_j'^{+} + Q_{ij} b_j'$$

*the coefficients $S^{ij}$ and $T_{ij}$ of the standard real form transform so that in the new variables $b_i'^{+}$, $b_i^{+}$ they become*

$$S = DS'D^\top, \quad T = D^{-1}T'(D^{-1})^\top,$$

*where*

$$D = P + Q, \quad D^{-1} = (P - Q)^\top.$$

The proof of this lemma consists in a direct verification based on the canonical relations:

$$[b_j^+, b_k^+] = [b_j, b_k] = [b_j'^{+}, b_k'^{+}] = [b_j', b_k'] = 0,$$
$$[b_j^+, b_k] = [b_j'^{+}, b_k'] = \delta_{jk}.$$

This lemma shows that the language of Bogolyubov transformations is equivalent in this case to the usual reduction of quadratic forms by a linear transformation.

Now we turn to fermions. In this case there is no classical analog, and diagonalization of quadratic forms was done in 1986 (S. P. Novikov) with the aim of constructing analogs of the Morse inequalities for critical points of vector fields.

**Lemma 7.13.** *Under real Bogolyubov transformations, the coefficients of the standard real representation of a fermion form are transformed by the rules*

$$R = O_+ R' O_-, \quad A = R_{ij}(a_i^+ + a_i)(a_j^+ + a_j),$$

*where $R = (R_{ij})$, $R' = (R'_{ij})$ with*

$$R_{ij} = u_{ij} + v_{ij}, \quad O_\pm = P \pm Q,$$

*$O_\pm$ being orthogonal transformations, $O_\pm \in \mathrm{O}(n)$.*

*Conversely, by these rules any pair of orthogonal matrices $O_\pm \in \mathrm{O}(n)$ determines a real Bogolyubov transformation of the fermion form.*

This lemma is also proved by an elementary direct substitution.

CONCLUSION. A real fermion quadratic form $A$ can always be diagonalized by a real Bogolyubov transformation (even such that $O_\pm \in \mathrm{SO}(n)$). In this case the matrix $R$ is diagonalized by a pair of orthogonal transformations $O_\pm$:

$$R' \to R = O_+ R' O_-.$$

The only invariants are the eigenvalues $\lambda_j^2$ of the matrix $RR^\top$ ("$s$-numbers"). The eigenvalues of the fermion form

$$A = \sum_{i,j} (u_{ij} a_i^+ a_j^+ + 2v_{ij} a_i^+ a_j - u_{ij} a_i a_j) + \sum_{i=1}^n v_{ii}$$

on a complete Fock space of fermions isomorphic to the complete exterior power

$$\Lambda^* \mathbb{R}^n = \sum_{j=0}^n \Lambda^j \mathbb{R}^n$$

have the form $(\pm\lambda_1 \pm \lambda_2 \pm \cdots \pm \lambda_n)$.

## 7.5. Polyvectors and the integral of anticommuting variables

**7.5.1. Anticommuting variables and superalgebras.** Let $b_1^+, \ldots, b_n^+$ be the boson creation operators, and let $b_1, \ldots, b_n$ be the dual to them annihilation operators introduced in Section 7.4.4. Consider their realization by operators acting on polynomials in $n$-dimensional Euclidean space with coordinates $x^1, \ldots, x^n$ (this realization is different from the one given in Section 7.4.4):

$$b_j^+(f(x)) = x^j f(x), \quad b_k(f(x)) = \frac{\partial f(x)}{\partial x^k}, \quad j, k = 1, \ldots, n.$$

In this case the function $f \equiv 1$ is the vacuum vector.

The creation operators for fermions cannot be realized in a similar manner, since they anticommute: $a_i a_j = -a_j a_i$. For their realization we must use the anticommuting variables. They are introduced as follows.

A *superspace* is defined to be any $\mathbb{Z}_2$-graded vector space

$$V = V_0 \oplus V_1.$$

The elements of $V_0$ are called even, and those of $V_1$ odd. This is indicated by the function $\alpha(a)$ defined on the elements of $V_0$ and $V_1$ as

$$\alpha(a) = \begin{cases} 0 & \text{for } a \in V_0, \\ 1 & \text{for } a \in V_1. \end{cases}$$

If an associative multiplication is given on a superspace such that

$$ab = (-1)^{\alpha(a)\alpha(b)} ba$$

on even and odd elements, then the superspace is called a *superalgebra*. This implies, in particular, that the multiplication is anticommutative on the odd elements.

EXAMPLES. 1. THE GRASSMANN ALGEBRA $\Lambda^*\mathbb{R}^n = \Lambda_0 \oplus \Lambda_1$ in $\mathbb{R}^n$. It consists of all exterior forms, which are skew-symmetric tensors of type $(0, k)$, with exterior product operation

$$\alpha, \beta \to \alpha \wedge \beta.$$

The space $\Lambda_0$ consists of all tensors of even rank, and the space $\Lambda_1$ consists of all tensors of odd rank.

2. THE ALGEBRA OF POLYVECTORS IN $\mathbb{R}^n$. A *polyvector* is a skew-symmetric tensor with superscripts only. The theory of polyvectors is similar to the theory of exterior forms.

1) The exterior product of polyvectors is specified as follows: if $R^{i_1 \dots i_p}$ and $S^{j_1 \dots j_q}$ are polyvectors of types $(p, 0)$ and $(q, 0)$, then

$$T^{k_1 \dots k_{p+q}} = \frac{1}{p!\,q!} \sum_{\sigma \in S_{p+q}} \operatorname{sgn}(\sigma) R^{\sigma(k_1)\dots\sigma(k_p)} S^{\sigma(k_{p+1})\dots\sigma(k_{p+q})}.$$

2) If $e_1, \dots, e_n$ is a basis of vectors at a point, then the polyvectors

$$e_{i_1} \wedge \dots \wedge e_{i_k}, \quad i_1 < \dots < i_k,$$

constitute a basis of polyvectors of type $(k, 0)$ at the point, and their contraction with the basis forms is

$$(e_{i_1} \wedge \dots \wedge e_{i_k}, e^{j_1} \wedge \dots \wedge e^{j_k}) = \delta_{i_1}^{j_1} \cdots \delta_{i_k}^{j_k}, \tag{7.15}$$

where $j_1 < \dots < j_k$ and $e^i = dx^i$.

3) At each point the polyvectors compose an exterior algebra $\Lambda^* V$ generated by the symbols $e_1 = \frac{\partial}{\partial x^1}, \dots, e_n = \frac{\partial}{\partial x^n}$, i.e., by the basis vectors of the tangent space $V$ at this point.

4) The following analog of Theorem 7.11 holds.

**Theorem 7.15.** *Under the changes of coordinates in the $n$-dimensional vector space, the polyvectors transform by the formula*

$$T^{1\ldots n} = J^{-1}\tilde{T}^{1\ldots n}, \tag{7.16}$$

*where $T = T^{1\ldots n}\frac{\partial}{\partial x^1} \wedge \cdots \wedge \frac{\partial}{\partial x^n} = \tilde{T}^{1\ldots n}\frac{\partial}{\partial y^1} \wedge \cdots \wedge \frac{\partial}{\partial y^n}$ is a polyvector of type $(n, 0)$ and $J = \det(\partial y^i/\partial x^j)$ is the Jacobian of the coordinate change.*

An example of a polyvector of type $(2,0)$ is given by the tensor of the Poisson structure $h^{jk}\frac{\partial}{\partial x^j} \wedge \frac{\partial}{\partial x^k}$.

We see that, relative to the exterior product, the Grassmann algebra of exterior forms at a point is generated by basis covectors $dx^1, \ldots, dx^n$, while the algebra of polyvectors is generated by the basis vectors $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$. Obviously, the algebra of polyvectors is a superalgebra relative to the exterior product.

Let $V = V_0 \oplus V_1$ be a superalgebra. Consider the linear space $V^{p|q}$ by all formal sums

$$\sum_{j=1}^{p} x_j e_j + \sum_{k=1}^{q} \xi_k e_{p+k},$$

where $e_1, \ldots, e_{p+q}$ are basis vectors in $V^{p|q}$. We regard the vectors $e_1, \ldots, e_p$ as even basis vectors, and $e_{p+1}, \ldots, e_{p+q}$ as odd basis vectors. The condition

$$x_j \in V_0,\ j = 1, \ldots, p, \quad \xi_k \in V_1,\ k = 1, \ldots, q,$$

specifies in $V^{p|q}$ a subspace $M$ endowed with two types of coordinates: the coordinates $x_1, \ldots, x_p$ commute relative to multiplication, while the coordinates $\xi_1, \ldots, \xi_q$ anticommute:

$$\xi_j \xi_k = -\xi_k \xi_j.$$

EXAMPLE. THE SPACE $\mathbb{R}^{p|q}$. Take the superalgebra $V$ to be the Grassmann algebra over the field $\mathbb{R}$ with $q$ generators. Then the space $M$ with $p$ even and $q$ odd coordinates is denoted by $\mathbb{R}^{p|q}$. By definition, smooth functions on this space are the functions

$$f(x, \xi) = \sum_{i_1 < \cdots < i_k} f_{i_1 \ldots i_k}(x_1, \ldots, x_p)\xi_{i_1} \wedge \cdots \wedge \xi_{i_k},$$

where $f_{i_1\ldots i_k}(x_1, \ldots, x_p)$ are real smooth functions of variables $x_1, \ldots, x_p$.

In the basis $(e_1, \ldots, e_{p+q})$, linear transformations of the space $V^{p|q}$ are specified by matrices of the form

$$T = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{7.17}$$

We will only consider linear transformations satisfying the condition

$$T(M) \subset M,$$

i.e., those that generate linear transformations of the space $M$. This means that the elements of the $p \times p$ matrix $A$ and $q \times q$ matrix $D$ are even, and those of the matrices $B$ and $C$ are odd.

The invertible matrices of this form constitute the group $\mathrm{GL}(p|q, V)$. We will give an important example of a subgroup of this group.

EXAMPLE. THE GROUP $\mathrm{OSp}(m|2n)$. Consider the following nondegenerate bilinear scalar product on the space $V^{m|2n}$:

$$(v, v') = \sum_{j=1}^{m} x_j x'_j + \sum_{k=1}^{2n} (\xi_k \xi'_{k+n} - \xi_{k+n} \xi'_k),$$

which is Euclidean in commuting coordinates and symplectic in anticommuting ones. The group of all transformations in $\mathrm{GL}(m|2n, V)$ preserving this scalar product is called the *orthogonal-symplectic group* and denoted by $\mathrm{OSp}(m|2n, V)$.

The groups $\mathrm{GL}(p|q, V)$ and $\mathrm{OSp}(p|q, V)$ are denoted by $\mathrm{GL}(p|q)$ and $\mathrm{OSp}(m|2n)$, respectively, when $V$ is a Grassmann algebra with $2n$ generators.

The trace of a transformation $T$ (*supertrace*) is defined by the formula

$$\operatorname{str} T = \operatorname{Tr} A - \operatorname{Tr} D$$

and takes values in the subalgebra $V_0$. In physics it is also denoted by $\operatorname{Tr}((-1)^F T)$.

**7.5.2. Integral of anticommuting variables.** Let $\xi_1, \dots, \xi_n$ be anticommuting variables and let $f(\xi_1, \dots, \xi_n)$ be a polynomial in the exterior algebra with generators $\xi_1, \dots, \xi_n$.

Following modern literature on quantum field theory, define the "integral of anticommuting variables",

$$\int \cdots \int_{\mathbb{R}^{0|n}} f(\xi_1, \dots, \xi_n)\, d\xi_1 \wedge \cdots \wedge d\xi_n,$$

by the following rules:

1) The integration is extended over the entire space, hence the integral is a functional on polynomials $f(\xi_1, \dots, \xi_n)$ in the exterior algebra.

2)

$$\int_{\mathbb{R}^{0|1}} \xi_j \, d\xi_j = 1, \quad \int_{\mathbb{R}^{0|1}} d\xi_j = 0.$$

3) The multiple integral is taken by iterated integration, e.g.,

$$\iint_{\mathbb{R}^{0|2}} f(\xi_1) \wedge g(\xi_2)\, d\xi_1 \wedge d\xi_2 = \left(\int_{\mathbb{R}^{0|1}} f(\xi_1)\, d\xi_1\right) \wedge \left(\int_{\mathbb{R}^{0|1}} g(\xi_2)\, d\xi_2\right).$$

4) Under linear changes of variables in the $\xi$-space, the expression $d\xi_1 \wedge \cdots \wedge d\xi_n$ is transformed by the rule

$$d\xi_1 \wedge \cdots \wedge d\xi_n = J^{-1}\, d\xi_1' \wedge \cdots \wedge d\xi_n',$$

where $J = \det(\partial \xi_i / \partial \xi_j')$.

Formulas (7.15) and (7.16) imply that integration so defined is the contraction of tensors in the Cartan formalism of exterior algebras for forms and polyvectors. Indeed, on associating $d\xi_j \to e_j$, $\xi_j \to e^j$, the expression $d\xi_1 \wedge \cdots \wedge d\xi_n$ reduces to the basis polyvector, which is a tensor of type $(n, 0)$, the leading term of the coefficient $f(\xi_1, \dots, \xi_n)$ reduces to a tensor of type $(0, n)$, and the "integral" becomes their ordinary contraction.

Example. The Gaussian integral of anticommuting variables. Let $f(\xi_1, \dots, \xi_n) = \exp\left(\frac{1}{2} a^{ij} \xi_i \wedge \xi_j\right)$, where $a^{ji} = -a^{ij}$. Then

$$\int \cdots \int_{\mathbb{R}^{0|n}} f(\xi_1, \dots, \xi_n)\, d\xi_1 \wedge \cdots \wedge d\xi_n = \sqrt{\det(a^{ij})}. \tag{7.18}$$

We can consider more general, not necessarily linear, changes of coordinates in the $\xi$-space requiring that they be $\mathbb{Z}_2$-graded. This means that they must take odd elements into odd elements. The partial derivatives of a function $f(\xi_1, \dots, \xi_n)$ of odd variables $\xi_1, \dots, \xi_n$ with values in the exterior algebra generated by these variables are defined by the formula

$$\frac{\partial f}{\partial \xi_i} = b_i,$$

where

$$f(\xi_1, \dots, \xi_n) = a_i(\xi_1, \dots, \hat{\xi}_i, \dots, \xi_n) + \xi_i \wedge b_i(\xi_1, \dots, \hat{\xi}_i, \dots, \xi_n).$$

Under $\mathbb{Z}_2$-graded changes of variables $(\xi_1, \dots, \xi_n) \to (\xi_1', \dots, \xi_n')$ the partial derivatives $\frac{\partial \xi_i}{\partial \xi_j'}$ are even elements of the exterior algebra; they commute with each other under multiplication, hence the Jacobian $\det\left(\frac{\partial \xi_i}{\partial \xi_j'}\right)$ is well defined.

Example. Consider the nonlinear change of coordinates:

$$\xi_1 = \xi_1' + \xi_1' \wedge \xi_2' \wedge \xi_3', \quad \xi_2 = \xi_2', \quad \xi_3 = \xi_3'.$$

The Jacobi matrix of this change is

$$\left(\frac{\partial \xi_i}{\partial \xi'_j}\right) = \begin{pmatrix} 1+\xi'_2 \wedge \xi'_3 & -\xi'_1 \wedge \xi'_3 & \xi'_1 \wedge \xi'_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

whence

$$J = \det\left(\frac{\partial \xi_i}{\partial \xi'_j}\right) = 1 + \xi'_2 \wedge \xi'_3, \quad J^{-1} = 1 - \xi'_2 \wedge \xi'_3.$$

The formula for the change of variables holds also for nonlinear coordinate changes:

$$\int_{\mathbb{R}^{0|n}} f(\xi(\xi'))J^{-1}\xi'_1 \wedge \cdots \wedge \xi'_n = \int_{\mathbb{R}^{0|n}} f(\xi)\xi_1 \wedge \cdots \wedge \xi_n, \tag{7.19}$$

where $J = \det(\partial\xi_i/\partial\xi'_j)$.

The derivation of formulas (7.18) and (7.19) is left as an exercise for the reader.

## Exercises to Chapter 7

1. Find the expression for the Euclidean metric on the plane in terms of polar coordinates.

2. Write down the metric tensor in $\mathbb{R}^3$ in cylindrical and spherical coordinates.

3. Show that if an invertible matrix $g_{ij}$ specifies a tensor of type $(0,2)$, the inverse matrix $(g^{ij})$, where $g^{ij}g_{jk} = \delta^i_k$, determines a tensor of type $(2,0)$.

4. What is the type of a tensor specifying a linear operator from the space of tensors of type $(k,s)$ into the space of tensors of type $(p,q)$?

5. Write down the gradient of a function, which is the vector field obtained as a composition of partial differentiation and raising indices, in polar, cylindrical, and spherical coordinates.

6. Let $\omega^j = a^j_i\,dx^i$. Show that

$$\omega^{i_1} \wedge \cdots \wedge \omega^{i_k} = J^{i_1\ldots i_k}_{j_1\ldots j_k}\,dx^{j_1} \wedge \cdots \wedge dx^{j_k},$$

where $J^{i_1\ldots i_k}_{j_1\ldots j_k}$ is the minor of the matrix $(a^j_i)$ placed in the intersection of the rows $i_1,\ldots,i_k$ and columns $j_1,\ldots,j_k$. For $k=n$ we have

$$\omega^1 \wedge \cdots \wedge \omega^n = \det(a^j_i)\,dx^1 \wedge \cdots \wedge dx^n.$$

7. Prove the formula

$$\sum_k a^j_{i_k}\varepsilon_{i_1\ldots i_{k-1}ji_{k+1}\ldots i_n} = \varepsilon_{i_1\ldots i_n}\operatorname{Tr}(a^j_i),$$

where $\operatorname{Tr}(a^j_i) = a^i_i$ is the trace of the matrix.

8. Express the vector product in $\mathbb{R}^3$ in terms of algebraic operations on tensors and operator $*$.

9. Give the classification of tensors of rank at most 4 relative to rotations in $\mathbb{R}^2$ (in $\mathbb{R}^3$) leaving the unit square (respectively, the unit cube in $\mathbb{R}^3$) unchanged.

10. Show that in Euclidean space $\mathbb{R}^n$ there are no tensors of odd rank invariant under rotations.

11. Prove formula (7.14):

$$*^2 = (-1)^{k(n-k)}$$

for the squared Hodge star-operator.

12. Show that the linear transformation of the superspace $V^{p|q}$ specified by the matrix (7.17) is invertible if and only if the matrices $A$ and $D$ are invertible.

13. Prove formula (7.18) for the Gaussian integral of anticommuting variables.

14. Prove formula (7.19) for the change of variables in the integral of anticommuting variables.

Chapter 8

# Tensor Fields in Analysis

## 8.1. Tensors of rank 2 in pseudo-Euclidean space

**8.1.1. Electromagnetic field.** The *electromagnetic field* $F_{ik}$ is specified by the *electric* and *magnetic* fields $\mathbf{E}$ and $\mathbf{H}$ by the formula

$$F = (F_{ik}) = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -H_3 & H_2 \\ -E_2 & H_3 & 0 & -H_1 \\ -E_3 & -H_2 & H_1 & 0 \end{pmatrix}. \tag{8.1}$$

Under rotations of three-dimensional Euclidean space $\mathbb{R}^3$, $\mathbf{E}$ and $\mathbf{H}$ are transformed as vectors (which allows us to write their components with subscripts). Under the Lorentz transformations

$$x^1 = \frac{\tilde{x}^1 + v\tilde{t}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x^2 = \tilde{x}^2, \quad x^3 = \tilde{x}^3, \quad t = \frac{t + \frac{v}{c^2}\tilde{x}^1}{\sqrt{1 - \frac{v^2}{c^2}}},$$

where $x^0 = ct$, $\tilde{x}^0 = c\tilde{t}$, the electric and magnetic fields are transformed by the following formulas:

$$E_1 = \tilde{E}_1, \quad E_2 = \frac{\tilde{E}_2 + \frac{v}{c}\tilde{H}_3}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad E_3 = \frac{\tilde{E}_3 - \frac{v}{c}\tilde{H}_2}{\sqrt{1 - \frac{v^2}{c^2}}},$$

$$H_1 = \tilde{H}_1, \quad H_2 = \frac{\tilde{H}_2 - \frac{v}{c}\tilde{E}_3}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad H_3 = \frac{\tilde{H}_3 + \frac{v}{c}\tilde{E}_2}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

This implies that the electromagnetic field $F_{jk}$ is a skew-symmetric tensor of rank 2 in the 4-dimensional Minkowski space $\mathbb{R}^{1,3}$. In terms of differential forms this statement is written as

$$
\begin{aligned}
F &= \sum_{j<k} F_{jk}\, dx^j \wedge dx^k \\
&= E_j\, dx^0 \wedge dx^j - H_1\, dx^2 \wedge dx^3 - H_2\, dx^3 \wedge dx^1 - H_3\, dx^1 \wedge dx^2.
\end{aligned} \tag{8.2}
$$

The *invariants of the field* $F_{jk}$ are defined to be the coefficients of the characteristic polynomial

$$P(\lambda) = \det(F_{jk} - \lambda g_{jk}).$$

Substituting formula (8.1) and the Minkowski metric $g_{jk}$ into this expression, we obtain

$$P(\lambda) = -\lambda^4 + (\mathbf{E}^2 - \mathbf{H}^2)\lambda^2 + \langle \mathbf{E}, \mathbf{H}\rangle^2.$$

At each point, the skew-symmetric tensors of rank 2 form a 6-dimensional linear space, on which the operator $*$ acts.

**Lemma 8.1.** *The operator* $*$ *acts on the field* (8.2) *by the formula*

$$*F = -H_i\, dx^0 \wedge dx^i - E_1\, dx^2 \wedge dx^3 - E_2\, dx^3 \wedge dx^1 - E_3\, dx^1 \wedge dx^2, \tag{8.3}$$

*i.e.*, $*(\mathbf{E}, \mathbf{H}) = (-\mathbf{H}, \mathbf{E})$.

**Proof.** By definition, the operator $*$ acts by the formula

$$(*F)_{jk} = \frac{1}{2}\varepsilon_{jklm}F^{lm},$$

where $F^{lm} = g^{lp}g^{mq}F_{pq}$ and $g_{jk}$ is the Minkowski metric. Hence

$$F^{0k} = -F_{0k} = -E_k, \quad F^{jk} = F_{jk} \quad \text{for} \quad j, k = 1, 2, 3.$$

Substituting this expression for $F^{jk}$ into the formula for the action of $*$, we obtain the lemma. □

**Corollary 8.1.** *The operator* $*$ *applied twice (squared) is* $-1$, *i.e.*, $*^2 = -1$.

If on a linear space, we have an operator such that its square equals $-1$, we can introduce complex coordinates in this space. Indeed, we identify this operator with multiplication by the imaginary unit $i = \sqrt{-1}$ and let

$$(a + bi)F = aF + b{*}F.$$

Since $i^2F = **F = -F$, this correspondence is well defined.

In the space $\mathbb{C}^3$ formed by all skew-symmetric tensors of rank 2, the field $F$ is written as

$$F = \mathbf{E} + i\mathbf{H}.$$

Indeed, $*F = -\mathbf{H} + i\mathbf{E} = iF$. Therefore, we will regard the fields $\mathbf{E}$ and $\mathbf{H}$ as the real and imaginary parts of the field $F$.

The operator $*$ is invariant under the action of the group $\mathrm{SO}(1,3)$, which preserves the Minkowski metric. Furthermore, the transformations in $\mathrm{SO}(1,3)$ preserve the characteristic polynomial of the field. This is due to the fact that they preserve the quadratic form

$$\langle F, F\rangle = -*(F\wedge *F + iF\wedge F) = -\frac{1}{2}\,(F_{jk}F^{kj} + i\varepsilon^{jklm}F_{jk}F_{lm}),$$

which in coordinates $z^k = E_k + iH_k$, $k = 1, 2, 3$, is written as

$$\langle F, F\rangle = (\mathbf{E} + i\mathbf{H})^2 = \sum_{k=1}^{3}(z^k)^2.$$

Hence we see that the transformations in $\mathrm{SO}(1,3)$ determine the linear transformations of $\mathbb{C}^3$ that preserve the scalar squares of complex vectors.

The group of complex-linear transformations of $\mathbb{C}^n$ that preserve the complex-valued scalar product $\langle z, w\rangle = \sum_{k=1}^{n} z^k w^k$ is denoted by $\mathrm{O}(n,\mathbb{C})$, not to be confused with the group $\mathrm{U}(n)$, which consists of transformations that preserve the Hermitian product $(z, w) = \sum_{k=1}^{n} z^k \bar{w}^k$.

Thus we obtained the homomorphism

$$\mathrm{SO}(1,3) \to \mathrm{O}(3,\mathbb{C}), \tag{8.4}$$

which is realized by action of Lorentz transformations on the complex space of fields $F_{jk}$. The invariants of Lorentz transformations are the real and imaginary parts of the scalar square of the field, $\mathrm{Re}\langle F, F\rangle = \mathbf{E}^2 - \mathbf{H}^2$ and $\mathrm{Im}\langle F, F\rangle = \langle \mathbf{E}, \mathbf{H}\rangle$.

Both groups $\mathrm{SO}(1,3)$ and $\mathrm{O}(3,\mathbb{C})$ have real dimension 6. It can be proved that the mapping (8.4) is an isomorphism of Lie groups, $\mathrm{SO}(1,3) \simeq \mathrm{O}(3,\mathbb{C})$.

**8.1.2. Reduction of skew-symmetric tensors to canonical form.** We have shown in 2.1.1 that any skew-symmetric form in a linear space can be reduced to the canonical form. But, of course, this fails to hold if we require that the reduction be done by a Lorentz transformation. In this case there are several "canonical" forms, which are described by the following theorem.

**Theorem 8.1.** 1. *Let $\langle F, F\rangle = \mathbf{E}^2 - \mathbf{H}^2 + 2i\langle \mathbf{E}, \mathbf{H}\rangle \neq 0$. In this case the field $F_{jk}$ is reduced by a Lorentz transformation to the form*

$$F = \begin{pmatrix} 0 & E & 0 & 0 \\ -E & 0 & 0 & 0 \\ 0 & 0 & 0 & -H \\ 0 & 0 & H & 0 \end{pmatrix}, \tag{8.5}$$

*where $E^2 - H^2 = \mathbf{E}^2 - \mathbf{H}^2$ and $EH = \langle \mathbf{E}, \mathbf{H}\rangle$.*

*Here the following two cases are possible:*

a) *if* $\langle \mathbf{E}, \mathbf{H} \rangle \neq 0$, *then* $EH \neq 0$;

b) *if* $\langle \mathbf{E}, \mathbf{H} \rangle = 0$ *and* $\mathbf{E}^2 - \mathbf{H}^2 \neq 0$, *then* $E \neq 0$, $H = 0$ *for* $\mathbf{E}^2 - \mathbf{H}^2 > 0$ *and* $E = 0$, $H \neq 0$ *for* $\mathbf{E}^2 - \mathbf{H}^2 < 0$.

2. *Let* $\langle F, F \rangle = 0$, *i.e.*, $\mathbf{E}^2 - \mathbf{H}^2 = \langle \mathbf{E}, \mathbf{H} \rangle = 0$. *Then the field* $F_{jk}$ *is reduced by a Lorentz transformation to the form*

$$F = \begin{pmatrix} 0 & E & 0 & 0 \\ -E & 0 & -E & 0 \\ 0 & E & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{8.6}$$

**Proof.** 1. Let $\langle F, F \rangle \neq 0$. The vector $F = \mathbf{E} + i\mathbf{H}$ can be written as $F = c\mathbf{n}$, where $c = \sqrt{\langle F, F \rangle}$ and $\langle \mathbf{n}, \mathbf{n} \rangle = 1$. Transform the vector $\mathbf{n}$ into the basis vector $e_1$ directed along the $x^1$-axis by a Lorentz transformation belonging to $\mathrm{O}(3, \mathbb{C}) = \mathrm{SO}(1, 3)$. The vectors $\mathbf{E}$ and $\mathbf{H}$ are transformed then into

$$\mathbf{E}' = (E', 0, 0), \quad \mathbf{H}' = (H', 0, 0),$$

where $E'H' = \langle \mathbf{E}, \mathbf{H} \rangle$. The field $F$ takes the form (8.5). If $\langle \mathbf{E}, \mathbf{H} \rangle \neq 0$, then the vectors $\mathbf{E}'$ and $\mathbf{H}'$ are parallel and both of them are not equal to zero. If $\langle \mathbf{E}, \mathbf{H} \rangle = 0$, then either $E' = 0$ or $H' = 0$ depending on the sign of the second invariant $\mathbf{E}^2 - \mathbf{H}^2 = E'^2 - H'^2$.

2. If $\mathbf{E}^2 - \mathbf{H}^2 = \langle \mathbf{E}, \mathbf{H} \rangle = 0$, then we transform the vector $\mathbf{E}$ into a vector parallel the $x^1$-axis by a real rotation in $\mathrm{O}(3, \mathbb{R}) \subset \mathrm{O}(3, \mathbb{C})$. The vector $\mathbf{H}$ goes then into a vector lying in the plane $x^1 = 0$, and it can be taken into a vector parallel the $x^3$-axis by a rotation of this plane. □

The physical meaning of this theorem is as follows: by transition to a new inertial coordinate system in the Minkowski space the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ may be reduced at any given point to one of the canonical forms (8.5) or (8.6).

Let $F_{jk}$ be a skew-symmetric (electromagnetic field) tensor in the Minkowski space $\mathbb{R}^{1,3}$. We construct a symmetric tensor $T_{jk}$ corresponding to $F_{jk}$ by setting

$$T_{jk} = \frac{1}{4\pi}\Big(-g^{lm} F_{jl} F_{km} + \frac{1}{4}\, F_{lm} F^{lm} g_{jk}\Big).$$

It is called the *energy-momentum tensor of the electromagnetic field.*

Its components have the form

$$T_{00} = \frac{\mathbf{E}^2 + \mathbf{H}^2}{8\pi}, \quad T_{0\alpha} = -\frac{1}{4\pi}\,[\mathbf{E}, \mathbf{H}]_\alpha,$$

$$T_{\alpha\beta} = \frac{1}{4\pi}\Big\{-E_\alpha E_\beta - H_\alpha H_\beta + \frac{1}{2}\,\delta_{\alpha\beta}(\mathbf{E}^2 + \mathbf{H}^2)\Big\},$$

where $\alpha, \beta = 1, 2, 3$. All of them have their own physical meaning: $T_{00} = W$ is called the *density of the energy* of the electromagnetic field, the vector $S_\alpha = -cT_{0\alpha}$ is called the *Poynting vector* (here $c$ is the velocity of light in vacuum), and the tensor $T_{\alpha\beta}$, $\alpha, \beta = 1, 2, 3$, is the *Maxwell stress tensor.*

We apply Theorem 8.1 to the tensor $T_{jk}$.

In the case of (8.5), where $\mathbf{E} = (E, 0, 0)$ and $\mathbf{H} = (H, 0, 0)$, we have

$$(T_{jk}) = \begin{pmatrix} W & & & 0 \\ & -W & & \\ & & W & \\ 0 & & & W \end{pmatrix}, \quad W = \frac{1}{8\pi}(E^2 + H^2),$$

i.e., the tensor $T_{jk}$ is diagonal.

In the case of (8.6), where $\mathbf{E} = (E, 0, 0), \mathbf{H} = (0, 0, H)$, and $E = H$, the tensor $T_{jk}$ is not reduced to the diagonal form and is equal to

$$(T_{jk}) = \begin{pmatrix} W & 0 & -W & 0 \\ 0 & 0 & 0 & 0 \\ -W & 0 & W & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad W = \frac{E^2}{4\pi} = \frac{H^2}{4\pi}.$$

Note that by definition, the trace of the energy-momentum tensor is always equal to zero:

$$T = g^{jk}T_{jk} = 0.$$

**8.1.3. Symmetric tensors.** In Euclidean space any symmetric matrix is reduced to the diagonal form by orthogonal transformations. Contrary to this, we have seen that in some cases the energy-momentum tensor of electromagnetic field cannot be diagonalized by means of Lorentz transformations.

Now we explore to which canonical form one can reduce a symmetric tensor of type $(0, 2)$ in pseudo-Euclidean space, taking as an example the two-dimensional space $\mathbb{R}^{1,1}$ with metric $g_{jk} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

The *eigenvalues of a (symmetric) tensor* $T_{jk}$ in a space with metric $g_{jk}$ are the solutions to the characteristic equation

$$P(\lambda) = \det(T_{jk} - \lambda g_{jk}) = 0.$$

Raising the subscripts, we see that they are eigenvalues of the operator $T^j_k = g^{jl}T_{lk}$. Indeed,

$$\det(T^j_k - \lambda\delta^j_k) = \det[g^{jl}(T_{lk} - \lambda g_{lk})] = \det g^{jl}\det(T_{lk} - \lambda g_{lk}).$$

The eigenvectors related to the eigenvalue $\lambda$ are the solutions to the equation

$$T^j_k\xi^k = \lambda\xi^j,$$

which, upon lowering superscripts, becomes

$$T_{jk}\xi^k = \lambda g_{jk}\xi^k.$$

**Theorem 8.2.** 1. *If the eigenvalues* $\lambda_0$ *and* $\lambda_1$ *of the tensor* $T_{jk}$ *are real and distinct, then the tensor can be reduced by a Lorentz transformation to the diagonal form*

$$\begin{pmatrix} \lambda_0 & 0 \\ 0 & -\lambda_1 \end{pmatrix}. \tag{8.7}$$

2. *If the eigenvalues are equal* ($\lambda_0 = \lambda_1 = \lambda$) *and hence real, then the tensor can be reduced by a Lorentz transformation to the form*

$$\begin{pmatrix} \lambda+\mu & -\mu \\ -\mu & -\lambda+\mu \end{pmatrix},$$

*where* $\mu$ *does not, in general, vanish and is not an invariant of the tensor.*

3. *If the eigenvalues are not real, then they are complex conjugate* ($\lambda_{0,1} = \alpha \pm i\beta$) *and the tensor is reduced to the form*

$$\begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix}. \tag{8.8}$$

**Proof.** 1. Let $Te_k = \lambda_k e_k$, $k = 0, 1$. Then the vectors $e_0$ and $e_1$ are orthogonal. Indeed,

$$\lambda_0\langle e_0, e_1\rangle = \langle Te_0, e_1\rangle = T_{jk}e_0^j e_1^k = \langle e_0, Te_1\rangle = \lambda_1\langle e_0, e_1\rangle,$$

and since $\lambda_0 \neq \lambda_1$, we have $\langle e_0, e_1\rangle = 0$. Take the coordinate axes directed along $e_0$ and $e_1$. Then the matrix $T_{jk}$ takes the form (8.7).

2. If there are two linearly independent eigenvectors corresponding to the same eigenvalue $\lambda$, then $T_k^j = \lambda\delta_k^j$ and the matrix $T_{jk}$ is reduced to the diagonal form. Suppose that there is a single eigenvector $\xi$, and let $L$ be the corresponding eigenspace. The orthogonal complement $L^\perp$ is invariant under the operator $T$: $\langle L, TL^\perp\rangle = \langle TL, L^\perp\rangle = \lambda\langle L, L^\perp\rangle = 0$. Therefore, if $L^\perp \neq L$, then $L^\perp$ contains another eigenvector. Hence $L = L^\perp$, which means that the vector $\xi$ is isotropic, $\langle \xi, \xi\rangle = 0$. Hence $\xi = (a, \pm a)$. We can normalize it by setting $a = 1$ and, applying a reflection if necessary, reduce $\xi$ to the form $\xi = (1, 1)$. Then the condition $(T_{jk} - \lambda g_{jk})\xi^k = 0$ becomes

$$T_{00} + T_{01} = \lambda, \quad T_{10} + T_{11} = -\lambda,$$

which, taking into account symmetry of $T_{jk}$, implies

$$T_{00} = \lambda + \mu, \quad T_{11} = -\lambda + \mu, \quad T_{01} = T_{10} = -\mu.$$

3. Let $e_{0,1} = u \pm iv$ be (complex) eigenvectors with eigenvalues $\lambda_{0,1}$. Since the eigenvalues are different, these eigenvectors are orthogonal,

$\langle u + iv, u - iv \rangle = 0$ (see part 1 of the proof). We normalize them by the conditions

$$\langle u + iv, u + iv \rangle = \langle u - iv, u - iv \rangle = 2.$$

Hence we see that

$$\langle u, v \rangle = 0, \quad \langle u, u \rangle = -\langle v, v \rangle, \quad \langle u, u \rangle - \langle v, v \rangle = 2.$$

Therefore, the vectors $u$ and $v$ form an orthonormal basis,

$$\langle u, u \rangle = -\langle v, v \rangle = 1, \quad \langle u, v \rangle = 0,$$

and the matrix $T_{jk}$ in this basis has the form (8.8). □

## 8.2. Behavior of tensors under mappings

**8.2.1. Action of mappings on tensors with superscripts.** Suppose $F\colon M^n \to N^k$ is a smooth mapping.

If $\xi$ is a tangent vector to $M^n$ at a point $x \in M^n$, then it is the velocity vector at $x$ of some curve $\gamma(t)$, $\dot\gamma(0) = \xi$, $\gamma(0) = x$.

Consider the image of this smooth curve under the mapping $F$. The velocity vector of the curve $F(\gamma(t))$ at $t = 0$ is tangent to the manifold $N^k$ at the point $F(x)$. Thus we obtain a mapping of tangent spaces,

$$F_*\colon T_x M^n \to T_{F(x)} N^k,$$

which by the chain rule is written in local coordinates as

$$F_* \xi^i = \frac{\partial y^i(x)}{\partial x^j} \frac{\partial x^j}{\partial t} = \frac{\partial y^i(x)}{\partial x^j} \xi^j,$$

where $x^1, \ldots, x^n$ are local coordinates on the manifold $M^n$ and $y^1, \ldots, y^k$ are local coordinates on $N^k$. It is seen from the representation in local coordinates that this mapping is linear and is specified by the Jacobi matrix of the mapping $F$. We have already mentioned it before; this mapping is called the *differential* of the mapping $F$ at the point $x$.

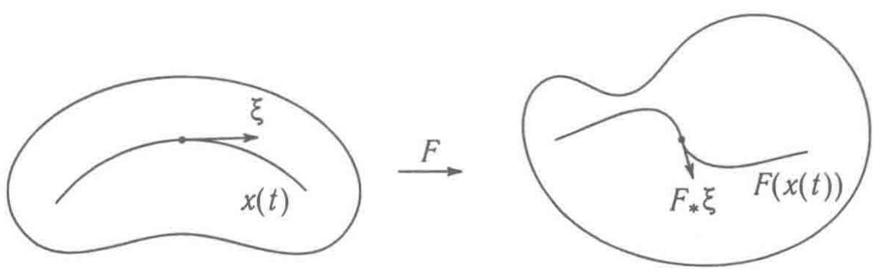


**Figure 8.1.** Differential of a smooth mapping.

In classical analysis we have a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$, and its differential is a linear mapping to the real line, $f_*\colon \mathbb{R}^n \to \mathbb{R}$, which determines

the mapping of tangent spaces specified by the gradient (the Jacobi matrix for a mapping into $\mathbb{R}$).

The tensor products of the basis vectors $e_{i_1} \otimes \cdots \otimes e_{i_l}$ form a basis in the space of tensors of type $(k,0)$; the differential is extended to these basis vectors in a natural way:

$$F_*(e_{i_1} \otimes \cdots \otimes e_{i_l}) = (F_* e_{i_1}) \otimes \cdots \otimes (F_* e_{i_l}).$$

Thus we obtain the following result.

**Theorem 8.3.** *A smooth mapping $F\colon M^n \to N^k$ generates a linear mapping of tensors of type $(l,0)$ at a point $x \in M^n$ into tensors of the same type at the point $F(x)$. This mapping is specified by the Jacobi matrix at the point $x$ by the formula:*

$$(8.9) \qquad \begin{aligned} (F_*T)^{i_1 \dots i_l}(F(x)) &= \frac{\partial y^{i_1}}{\partial x^{j_1}} \cdots \frac{\partial y^{i_l}}{\partial x^{j_l}} T^{j_1 \dots j_l}(x), \\ F(x) &= (y^1(x), \dots, y^k(x)), \end{aligned}$$

*and for vector fields (tensors of type $(1,0)$) coincides with the differential.*

If the manifolds $M^n$ and $N^k$ have the same dimension and the Jacobi matrix at a point $x \in M^n$ is invertible, then by the inverse function theorem the mapping $F$ is a local diffeomorphism in a neighborhood of this point, the mapping $F_*$ is invertible, and hence the inverse mapping $F_*^{-1}$ is defined for tensors of type $(l,0)$ with any $l \geq 1$.

**8.2.2. Restriction of tensors with subscripts.** Now we consider the action of a smooth mapping $F\colon M^n \to N^k$ on tensors of type $(0,l)$, i.e., tensors with subscripts.

Let $x^1, \dots, x^n$ be local coordinates in a neighborhood of a point $x \in M^n$ and $y^1, \dots, y^k$ local coordinates in a neighborhood of the point $F(x) \in N^k$. Any tensor of type $(0,l)$ at the point $F(x)$ is decomposed into a sum of the form

$$T = T_{i_1 \dots i_l}\, dy^{i_1} \otimes \cdots \otimes dy^{i_l}.$$

We restrict this tensor to the image of the mapping $F$. Then, setting $y^i = y^i(x)$, $i = 1, \dots, k$, we obtain a tensor at the point $x$,

$$\begin{aligned} &T_{i_1 \dots i_l}(F(x))\, dy^{i_1}(x) \otimes \cdots \otimes dy^{i_l}(x) \\ &\quad = T_{i_1 \dots i_l}(F(x)) \Big( \frac{\partial y^{i_1}(x)}{\partial x^{j_1}}\, dx^{j_1} \Big) \otimes \cdots \otimes \Big( \frac{\partial y^{i_l}(x)}{\partial x^{j_l}}\, dx^{j_l} \Big) \\ &\quad = \Big( T_{i_1 \dots i_l}(F(x)) \frac{\partial y^{i_1}(x)}{\partial x^{j_1}} \cdots \frac{\partial y^{i_l}(x)}{\partial x^{j_l}} \Big)\, dx^{j_1} \otimes \cdots \otimes dx^{j_l}. \end{aligned}$$

Thus we have proved the following theorem.

**Theorem 8.4.** *A smooth mapping $F\colon M^n \to N^k$ generates a linear mapping of tensors of type $(0,l)$ at the point $F(x) \in N^k$ into tensors of the same type at the point $x$. This mapping is given by the formula*

(8.10)
$$(F^*T)_{i_1\dots i_l}(x) = \frac{\partial y^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial y^{j_l}}{\partial x^{i_l}} T_{j_1\dots j_l}(F(x)), \quad F(x) = (y^1(x), \dots, y^k(x)).$$

The operation $F^*$ is called the *restriction of a tensor* (to the image of the mapping $F$).

EXAMPLES. 1. COVECTORS. First of all, we point out that a smooth mapping $F\colon M^n \to N^k$ acts on tensors with subscripts and superscripts in opposite directions. For example, for vectors and covectors we have, respectively,

$$F_*\colon T_xM^n \to T_{F(x)}N^k, \quad F^*\colon T^*_{F(x)}N^k \to T^*_xM^n.$$

Formulas (8.9) and (8.10) imply that these mappings are specified by conjugate matrices and preserve the action of covectors on vectors.

**Corollary 8.2.** *Let $\xi$ be a vector at a point $x \in M^n$, and $\eta$ a covector at the point $F(x) \in N^k$. Then*

$$\langle F^*\eta, \xi\rangle = \langle \eta, F_*\xi\rangle.$$

2. RIEMANNIAN METRIC. Let $g_{ij}$ be a Riemannian metric in the space $N^k$ with local coordinates $y^1, \dots, y^k$, and let $F\colon M^n \to N^k$ be an imbedding of an $n$-dimensional surface with local coordinates $x^1, \dots, x^n$:

$$y^i = y^i(x^1, \dots, x^n), \quad i = 1, \dots, k.$$

By means of restriction operation we obtain the metric $\hat{g}_{ij}$ on the imbedded surface $M^n$,

$$\hat{g}_{ij} = g_{kl}\frac{\partial y^k}{\partial x^i}\frac{\partial y^l}{\partial x^j}.$$

This metric is called the *induced metric* (it is induced by the imbedding $F$). When $M^2$ is a regular surface in $N^k = \mathbb{R}^3$, it is also called the *first fundamental form of the surface* (see 3.2.1).

3. SKEW-SYMMETRIC TENSORS. Consider the restriction of an $n$-form to a submanifold of dimension $n$.

**Theorem 8.5.** *Let $M^n$ be an $n$-dimensional submanifold of $N^k$ specified by functions $y^i = y^i(x^1, \dots, x^n)$, $i = 1, \dots, k$, and let*

$$\omega = \sum_{i_1 < \dots < i_n} T_{i_1\dots i_n}\, dy^{i_1} \wedge \dots \wedge dy^{i_n}$$

*be an $n$-form on $N^k$. The restriction of the $n$-form $\omega$ to the $n$-dimensional submanifold $M^n$ is*

$$\sum_{i_1<\cdots<i_n} T_{i_1\dots i_n}\,dy^{i_1}\wedge\cdots\wedge dy^{i_n} = \Big(\sum_{i_1<\cdots<i_n} J^{i_1\dots i_n} T_{i_1\dots i_n}\Big)\,dx^1\wedge\cdots\wedge dx^n,$$

*where $J^{i_1\dots i_n}$ is the $n\times n$ minor of the matrix $\left(\frac{\partial y^i}{\partial x^j}\right)$ that is composed from the $(i_1,\dots,i_n)$th columns.*

**Proof.** Since

$$(F^*T)_{1\dots n} = T_{i_1\dots i_n}\frac{\partial y^{i_1}}{\partial x^1}\cdots\frac{\partial y^{i_n}}{\partial x^n}$$

and the tensor $T$ is skew-symmetric, we have

$$\begin{aligned} T_{i_1\dots i_n}\frac{\partial y^{i_1}}{\partial x^1}\cdots\frac{\partial y^{i_n}}{\partial x^n} &= \sum_{i_1<\cdots<i_n} T_{i_1\dots i_n}\Big(\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\frac{\partial y^{i_{\sigma(1)}}}{\partial x^1}\cdots\frac{\partial y^{i_{\sigma(n)}}}{\partial x^n}\Big)\\ &= \sum_{i_1<\cdots<i_n} T_{i_1\dots i_n}J^{i_1\dots i_n}. \end{aligned}$$

Hence the theorem. □

**8.2.3. The Gauss map.** As an example, we consider the action of the Gauss map on the forms.

Let $N^k$ be a submanifold in Euclidean space $\mathbb{R}^n$. At each point $x\in N^k$ consider the tangent space, which is realized as a $k$-dimensional plane in $\mathbb{R}^n$. To this plane corresponds a point of the Grassmann manifold $G_{n,k}$. Define the mapping

$$\psi\colon N^k\to G_{n,k},\quad \psi(x)=T_xN^k\in G_{n,k}.$$

It is referred to as the *Gauss map.* If the submanifold $N^k$ is oriented, then to each of its points we can assign the oriented tangent space to obtain the mapping

$$\psi\colon N^k\to\tilde{G}_{n,k}.$$

Consider the case of a hypersurface $M^{n-1}\subset\mathbb{R}^n$ in more detail. The hypersurface is locally specified as the set of zeros of a smooth function,

$$F(x^1,\dots,x^n)=0,\quad \operatorname{grad}F\neq 0,$$

where $x^1,\dots,x^n$ are Euclidean coordinates.

On the hypersurface the "curvature form"

$$K\,d\sigma = K\sqrt{g}\,dy^1\wedge\cdots\wedge dy^{n-1}$$

is defined, where $K$ is the curvature, which for $n=2$ is the curvature of the curve on the plane, and for $n=3$, the Gaussian curvature. We do not give the definition in the general case.

The Grassmann manifold $\tilde{G}_{n,n-1}$ is diffeomorphic to the sphere $S^{n-1}$ in $\mathbb{R}^n$ (see 6.1.5). Indeed, to each oriented $(n-1)$-dimensional plane the Gauss map assigns a unique unit normal vector $n$ to the plane which together with a positively oriented basis $\xi_1, \ldots, \xi_{n-1}$ of the plane forms a positively oriented basis $(\xi_1, \ldots, \xi_{n-1}, n)$ in $\mathbb{R}^n$.

On the sphere $S^{n-1}$ specified by the equation $\sum (x^k)^2 = 1$ we have the volume element $\Omega_{n-1}$, which is rotationally invariant and which for $n = 2, 3$ has the form

$$\begin{aligned} \Omega_1 &= d\varphi && \text{for } n = 2, \\ \Omega_2 &= \sin\theta \, d\theta \, d\varphi && \text{for } n = 3. \end{aligned}$$

**Theorem 8.6.** *The Gauss map of the hypersurface,*

$$\psi \colon M^{n-1} \to S^{n-1},$$

*acts on the volume form by the formula*

$$K \, d\sigma = \psi^*(\Omega_{n-1}),$$

*where $d\sigma = \sqrt{g} \, dy^1 \wedge \cdots \wedge dy^{n-1}$ is the element of $(n-1)$-dimensional volume in local coordinates $y^1, \ldots, y^{n-1}$ on the hypersurface $M^{n-1}$.*

*For $n = 2, 3$ we have*

$$\begin{aligned} &K \, dl = \psi^*(d\varphi) \quad \text{for } n = 2, \\ &K \sqrt{g} \, dy^1 \wedge dy^2 = \psi^*(\Omega_2), \quad \Omega_2 = \sin\theta \, d\theta \, d\varphi, \quad \text{for } n = 3. \end{aligned}$$

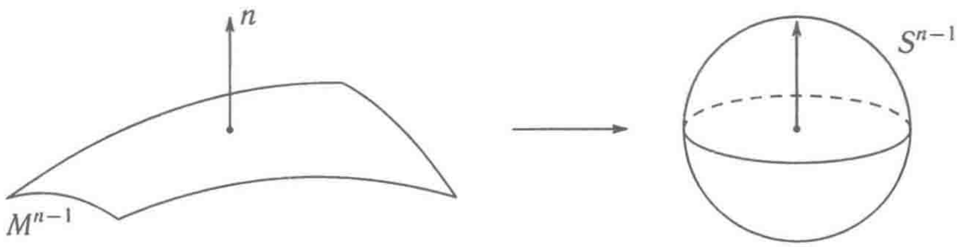


**Figure 8.2.** The Gauss map of a hypersurface.

**Proof.** We will give the proof for $n = 3$. The proof for the other cases is similar. In $\mathbb{R}^3$, take Euclidean coordinates with origin at a point $P \in M^2$ such that the $z$-axis is orthogonal to the surface at the point $P$, while the $x$- and $y$-axes are tangent to the surface. Then in a neighborhood of $P$ the surface is specified by an equation $z = f(x, y)$, where $df|_P = 0$.

Moreover, on the sphere $S^2$ we take coordinates of the same kind relative to the point $\psi(P) = Q$, namely, with $\tilde{z}$-axis orthogonal to the sphere and $\tilde{x}$- and $\tilde{y}$-axes tangent to it at the point $Q$.

In these coordinates, at the point $P$, we have

$$f_x = f_y = 0, \quad K = \det\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}, \quad g_{ij} = \delta_{ij}.$$

In a neighborhood of the point $Q$ the sphere is specified by the equation

$$\tilde{z} = \sqrt{1 - \tilde{x}^2 - \tilde{y}^2},$$

and the volume form at the point $Q$ is

$$\Omega_2 = d\tilde{x} \wedge d\tilde{y}.$$

The coordinates of the normal vector at a point of the surface lying near the point $P$ are

$$n = \left(\frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}}, \quad \frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}}, \quad \frac{-1}{\sqrt{1 + f_x^2 + f_y^2}}\right),$$

and $n = (0, 0, -1)$ at the point $P$. Therefore, in a neighborhood of $P$ the Gauss map $\psi$ has the form

$$(x, y) \to (\tilde{x}, \tilde{y}): \quad \tilde{x} = \frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}}, \quad \tilde{y} = \frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}}.$$

Hence

$$\psi^*(\Omega_2)|_P = \left(\frac{\partial \tilde{x}}{\partial x}\frac{\partial \tilde{y}}{\partial y} - \frac{\partial \tilde{x}}{\partial y}\frac{\partial \tilde{y}}{\partial x}\right)\Big|_P dx \wedge dy = J\, dx \wedge dy,$$

where $J$ is the Jacobian of the mapping $\psi$ at the point $P$.

Since $f_x = f_y = 0$ at $P$, we have $|g| = 1$ and $J = f_{xx}f_{yy} - f_{xy}f_{yx} = K$ at this point.

Thus we conclude that in the chosen coordinate system the following formula holds at the point $P$:

$$K\sqrt{|g|}\, dx \wedge dy = \psi^*(\Omega_2).$$

Since both sides of this equality are tensors, they coincide in all coordinate systems. Hence the theorem. □

## 8.3. Vector fields

**8.3.1. Integral curves.** We recall that an ordinary differential equation (ODE) is an equation of the form

$$\dot{x} = v(x), \tag{8.11}$$

specified in a domain $U \subset \mathbb{R}^n$ by a vector field $v(x)$. In the coordinate form,

$$\dot{x}^i = v^i(x), \quad i = 1, \dots, n.$$

More precisely, this formula specifies an autonomous ODE (i.e., such that the right-hand side does not depend on time). But any nonautonomous ODE

$$\dot{x} = v(x, t), \quad t \in I \subset \mathbb{R},$$

can be reduced to the autonomous form in the extended space $U \times I$ with coordinates $y = (x, t)$:

$$\dot{y} = (\dot{x}, \dot{t}) = (v(x, t), 1).$$

The solutions $x(t) = (x^1(t), \dots, x^n(t))$ to the equation (8.11) are called its *integral curves*.

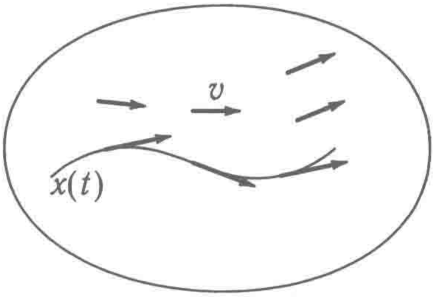


**Figure 8.3.** An integral curve.

By the theorem on the existence and uniqueness of the solution to an ODE we have the following statement.

*If $v(x)$ is a smooth vector field, then for any point $x_0 \in U$ there exists a neighborhood $V$ of this point and a positive $\varepsilon$ (depending on $x_0$) such that for all $x' \in V$ and $t \in [-\varepsilon, \varepsilon]$ there exists a unique solution to equation* (8.11) *with initial conditions*

$$x(0) = x' \in V.$$

*This solution is a smooth function of the initial condition $x'$.*

It follows from this theorem that there exists (locally) the translation mapping along integral curves,

$$\varphi_t \colon U \to U,$$

which associates with a point $x$ its translation along the integral curve for time $t$. To this end, consider the solution $\varphi(s)$ of the ODE

$$\frac{d\varphi}{ds} = v(\varphi(s))$$

with initial condition $\varphi(0) = x$, and set

$$\varphi_t(x) = \varphi(t).$$

Let $v$ be a smooth vector field, which in any coordinate neighborhood of a smooth manifold $M^n$ specifies an equation of the form (8.11). Then we obtain an ODE on the smooth manifold $M^n$. The theorem on the existence and uniqueness of a solution is of local nature and therefore holds also in this case. The existence of a global solution is provided by the following lemma.

**Lemma 8.2.** *If $v$ is a smooth vector field on a compact smooth manifold $M^n$, then for any point $x_0 \in M^n$ there exists a unique solution to equation* (8.11) *with initial condition $x(0) = x_0$ for all $t \in \mathbb{R}$, which is a smooth function of the initial condition $x_0$.*

**Proof.** For any point $x \in M^n$, find a neighborhood $V_x$ such that for $t \in [-\varepsilon_x, \varepsilon_x]$ there exists a unique solution to equation (8.11) with initial conditions in $V_x$. From the open covering $\{V_x\}$ we can select a finite subcovering $\{V_i = V_{x_i}\}$, since the manifold $M^n$ is compact. Let $\varepsilon = \min_i \varepsilon_{x_i}$. We see that for any point $x$ there is a unique translation $\varphi_u$ for $u \in [-\varepsilon, \varepsilon]$. Define the translation $\varphi_t$ as the iteration

$$\varphi_t = \underbrace{\varphi_\varepsilon \circ \cdots \circ \varphi_\varepsilon}_{n} \circ \varphi_s, \quad t = n\varepsilon + s, \quad 0 \le s < \varepsilon.$$

Then it is uniquely defined and determines a smooth mapping $\varphi_t \colon M^n \to M^n$. Hence the lemma. □

This proof carries over almost unchanged to the case where $v$ is a smooth vector field on a manifold $M^n$ and the closure of its support (the set of points where $v \neq 0$) is compact. For $M^n = \mathbb{R}^n$ this means that the vector field vanishes outside a bounded domain.

To each vector field corresponds a one-parameter *local group* of diffeomorphisms $\varphi_t$. This means that for small $t$ (possibly dependent on domains of the space) there are diffeomorphisms $\varphi_t$ satisfying the axioms of group theory,

$$\varphi_{s+t} = \varphi_s \circ \varphi_t, \quad (\varphi_t)^{-1} = \varphi_{-t}$$

whenever both sides of these equalities are (locally) defined. The diffeomorphisms $\varphi_t$ are translations in time $t$ along the integral curves of the vector field.

The Taylor expansion of the solution to ODE yields

$$\varphi_t^i(x) = x^i + tv^i(x) + o(t), \quad x = (x^1, \ldots, x^n).$$

Therefore, the Jacobi matrix of the mapping $\varphi_t$ equals

$$\frac{\partial \varphi_t^i}{\partial x^j} = \delta_j^i + t\frac{\partial v^i}{\partial x^j} + o(t). \tag{8.12}$$

Obviously, there exists also the converse correspondence: to each local one-parameter group of diffeomorphisms $\varphi_t$ corresponds a vector field $v$ tangent to its trajectory,

$$v(x) = \frac{d\varphi_t(x)}{dt}.$$

It is called the *field of velocities.*

EXAMPLE. On the plane with coordinates $(x, y)$, consider the linear vector field

$$\xi(x, y) = (-y, x).$$

The group $\varphi_t$ consists of rotations of the plane about the origin clockwise through the angle $t$:

$$\varphi_t = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

If the vector field is given on a compact manifold, then Lemma 8.2 implies the global correspondence:

*To each smooth vector field on a compact manifold $M^n$ corresponds a one-parameter group of diffeomorphisms $\varphi_t$, $t \in \mathbb{R}$, defining a smooth action of the group $\mathbb{R}$.*

When the group of diffeomorphisms corresponding to the vector field on a manifold $M^n$ is isomorphic to $\mathbb{R}$, one says that there is a smooth *flow* on the manifold,

$$\varphi_t \colon M^n \to M^n, \quad \varphi_{s+t} = \varphi_s \circ \varphi_t, \quad t \in \mathbb{R},$$

with $\varphi_0$ being the identity mapping.

**8.3.2. Lie algebras of vector fields.** Let $\varphi_t$ be a one-parameter group of diffeomorphisms of the manifold $M^n$ generated by the vector field $\xi$.

We restrict a smooth function $f \colon M^n \to \mathbb{R}$ to the integral curve of the field $\xi$ passing through the point $x_0$ at time $t = 0$. Then we obtain a function of $t$ on the integral curve and consider its derivative with respect to $t$ at $t = 0$. As a result, we obtain the *directional derivative of $f$ at the point $x_0$ in the direction of the field $\xi$*:

$$\partial_\xi f(x_0) = \frac{df(\varphi_t(x_0))}{dt}\bigg|_{t=0} = \frac{df(x_0 + t\xi(x_0) + o(t))}{dt}\bigg|_{t=0} = \xi^i \frac{\partial f}{\partial x^i}.$$

In this way, to each vector field corresponds a first-order differential operator,

$$\xi \to \partial_\xi = \xi^i \frac{\partial}{\partial x^i}.$$

It can be easily verified that this operator defines differentiation of the algebra of smooth functions, i.e., it is linear and satisfies the Leibniz identity

$$\partial_\xi(fg) = (\partial_\xi f)g + f(\partial_\xi g).$$

**Theorem 8.7.** *Let $\xi$ and $\eta$ be smooth vector fields. Then the commutator of the operators $\partial_\xi$ and $\partial_\eta$ is the first-order differential operator*

$$\partial_{[\xi,\eta]} = [\partial_\xi, \partial_\eta] = \partial_\xi\partial_\eta - \partial_\eta\partial_\xi,$$

*where the vector field $[\xi,\eta]$ has the form*

$$[\xi,\eta]^i = \xi^j \frac{\partial \eta^i}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j}. \tag{8.13}$$

**Proof.** This is shown by a direct calculation:

$$\begin{aligned}\partial_\xi(\partial_\eta f) - \partial_\eta(\partial_\xi f) &= \partial_\xi\Big(\eta^i \frac{\partial f}{\partial x^i}\Big) - \partial_\eta\Big(\xi^i \frac{\partial f}{\partial x^i}\Big) \\ &= \xi^j \frac{\partial \eta^i}{\partial x^j}\frac{\partial f}{\partial x^i} + \xi^j\eta^i \frac{\partial^2 f}{\partial x^i \partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j}\frac{\partial f}{\partial x^i} - \eta^j\xi^i \frac{\partial^2 f}{\partial x^i \partial x^j} \\ &= \Big(\xi^j \frac{\partial \eta^i}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j}\Big)\frac{\partial f}{\partial x^i}.\end{aligned}$$

□

The vector field $[\xi,\eta]$ of the form (8.13) is called the *commutator of the vector fields* $\xi$ and $\eta$.

Since operators of the form $\partial_\xi$, obviously, satisfy the Jacobi identity

$$[[\partial_\xi, \partial_\eta], \partial_\tau] + [[\partial_\eta, \partial_\tau], \partial_\xi] + [[\partial_\tau, \partial_\xi], \partial_\eta] = 0$$

relative to the usual commutator $[\partial_\xi, \partial_\eta] = \partial_\xi\partial_\eta - \partial_\eta\partial_\xi$, Theorem 8.7 implies the following result.

**Corollary 8.3.** *Smooth vector fields satisfy the Jacobi identity*

$$[[\xi,\eta],\tau] + [[\eta,\tau],\xi] + [[\tau,\xi],\eta] = 0,$$

*and hence they form a Lie algebra relative to the commutator.*

The Lie algebra of smooth vector fields is, obviously, infinite-dimensional and contains many subalgebras originating in geometry. In particular, the following theorem holds.

**Theorem 8.8.** *If $N^k$ is a submanifold of the manifold $M^n$, then smooth vector fields tangent to $N^k$ form a Lie subalgebra in the Lie algebra of smooth vector fields on $M^n$.*

**Proof.** Since the statement is of local nature, it suffices to prove it for a domain where the submanifold is specified by the equations $x^1 = \cdots = x^{n-k} = 0$ in appropriate local coordinates. The condition that $\xi$ is a tangent field is $\xi^1 = \cdots = \xi^{n-k} = 0$. Let $\xi$ and $\eta$ be vector fields tangent to $N^k$. This means, in particular, that

$$\frac{\partial \xi^i}{\partial x^j} = \frac{\partial \eta^i}{\partial x^j} = 0, \quad i = 1, \ldots, n-k, \quad j = n-k+1, \ldots, n,$$

at the points of $N^k$, since $x^{n-k+1}, \dots, x^n$ are local coordinates in $N^k$. Therefore, at a point of $N^k$, for $i = 1, \dots, n-k$, we have

$$[\xi, \eta]^i = \xi^j \frac{\partial \eta^i}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j} = \sum_{j=1}^{n-k} \Big( \xi^j \frac{\partial \eta^i}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j} \Big),$$

and the right-hand side is equal to zero, since $\xi^i = \eta^i = 0$ for $i = 1, \dots, n-k$ on the submanifold $N^k$. Thus we conclude that $[\xi, \eta]^1 = \cdots = [\xi, \eta]^{n-k} = 0$ on $N^k$. This means that the commutator of tangent vector fields is also tangent to the submanifold. □

**Corollary 8.4.** *If nonzero vector fields $\xi_1, \dots, \xi_n$ in a domain of $\mathbb{R}^n$ are tangent to the coordinate lines $x^1, \dots, x^n$, then their commutators have the form*

$$[\xi_i, \xi_j] = a_{ij} \xi_i + b_{ij} \xi_j, \quad i, j = 1, \dots, n$$

*(the rule of summation over repeated indices does not apply here).*

**Proof.** Any pair of vector fields $\xi_i$ and $\xi_j$ at any point $(x_0^1, \dots, x_0^k)$ is tangent to a manifold of the form $x^k = x_0^k$, $k \neq i, j$. Therefore, their commutator must also be tangent to this manifold. Hence the corollary. □

**8.3.3. Linear vector fields.** Let $A = (a_j^i)$ be an $n \times n$ matrix. It determines a vector field $T_A$ in $\mathbb{R}^n$ by the formula

$$\dot{x} = T_A(x) = -Ax,$$

or, in the coordinate form,

$$\dot{x}^i = -a_j^i x^j.$$

Such a field $T_A$ is called a *linear vector field*, since it depends linearly on the coordinates $x^1, \dots, x^n$.

**Theorem 8.9.** *The commutator of linear vector fields $T_A$ and $T_B$ is equal to*

$$[T_A, T_B] = T_{[A,B]}.$$

*Therefore, the linear vector fields form a finite-dimensional Lie subalgebra of the Lie algebra of smooth vector fields in $\mathbb{R}^n$. This subalgebra is isomorphic to the algebra $\mathfrak{gl}(n)$ formed by all $n \times n$ matrices.*

**Proof.** This is shown by a direct calculation:

$$\begin{aligned} [T_A, T_B]^i &= (-a_k^j x^k) \frac{\partial(-b_l^i x^l)}{\partial x^j} - (-b_k^j x^k) \frac{\partial(-a_l^i x^l)}{\partial x^j} \\ &= a_k^j x^k b_j^i - b_k^j x^k a_j^i = -(a_j^i b_k^j - b_j^i a_k^j) x^k = T_{[A,B]}^i. \end{aligned}$$

Hence the theorem. □

The following theorem shows that the integral curves of linear vector fields are found very easily.

**Theorem 8.10.** *The integral curve of a linear vector field $T_A$ that passes through a point $x_0$ at time $t = 0$ has the form*

$$x(t) = e^{-tA}x_0.$$

**Proof.** Recall that the exponential function of a matrix $A$ is defined as the convergent series

$$e^A = 1 + A + \frac{1}{2}A^2 + \cdots + \frac{1}{n!}A^n + \cdots$$

Differentiating the series for $e^{-tA}$ with respect to $t$ termwise, we obtain the series converging to $-Ae^{-tA}$. This implies that

$$\frac{d}{dt}(e^{-tA}x_0) = -A(e^{tA}x_0);$$

hence the vector-function $F(t) = e^{-tA}x_0$ satisfies the differential equation $\dot{x} = -Ax$ and $F(0) = x_0$. As is well known, the solution to this equation with a given initial condition is unique. Hence the theorem. □

We see that the one-parameter group of diffeomorphisms $\varphi_t$ generated by a linear vector field $T_A$ acts on $\mathbb{R}^n$ linearly and coincides with the one-parameter subgroup of the group $\mathrm{GL}(n)$ generated by the tangent matrix $-A \in \mathfrak{gl}(n)$.

Theorem 8.9 implies the following result.

**Corollary 8.5.** *Any matrix Lie algebra $\mathfrak{g} \subset \mathfrak{gl}(n)$ generates a Lie subalgebra of the algebra of smooth vector fields by the correspondence*

$$A \to T_A, \quad A \in \mathfrak{g}.$$

Let $G \subset \mathrm{GL}(n)$ be a matrix Lie group and $\mathfrak{g}$ its Lie algebra with a basis $e_1, \ldots, e_k$, and let $T_{e_1}, \ldots, T_{e_n}$ be the corresponding linear vector fields in $\mathbb{R}^n$. Then the differentiation operators in the direction of these fields are called the *generators of the (matrix) group $G$*. For example, consider the group $\mathrm{SO}(3)$; its Lie algebra has a basis

$$e_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with commutation relations $[e_i, e_j] = \varepsilon_{ijk}e_k$. The vector fields corresponding to this basis are

$$L_x = (0, z, -y), \quad L_y = (-z, 0, x), \quad L_z = (y, -x, 0).$$

They generate rotations about the $x$-, $y$-, and $z$-axes, respectively. The generators of the group of rotations SO(3) are denoted by the same letters:

$$L_x = z\frac{\partial}{\partial y} - y\frac{\partial}{\partial z}, \quad L_y = x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}, \quad L_z = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}.$$

**8.3.4. Exponential function of a vector field.** The generators of a matrix group generate its regular representation by means of the exponential mapping. We explain what it means.

Suppose we are given the action of a group $G$ on a space $X$. Consider the linear space $C(X)$ of functions on this space. The action of the group $G$ generates a *regular representation* of $G$ in the linear space $C(X)$ by the formula

$$f \xrightarrow{g} gf, \quad gf(x) = f(gx), \qquad g \in G, \quad f \in C(X), \quad x \in X.$$

Let $G$ be a one-parameter group of diffeomorphisms $\varphi_t \colon \mathbb{R}^n \to \mathbb{R}^n$. If it is generated by a constant vector field $\xi$, then

$$\varphi_t f(x) = f(x + t\xi).$$

If this group is generated by a linear vector field $T_A$, then

$$\varphi_t f(x) = f(e^{-tA}x).$$

Similarly to the case of matrix groups acting on finite-dimensional linear spaces, we represent $\varphi_t$ as an exponential function of the "tangent vector" at the identity to the group of all diffeomorphisms. This group is infinite-dimensional, and the "tangent vectors" are vector fields on a manifold. Thus we arrive at the formal definition of the *exponential function of the vector field* $\xi$ on $\mathbb{R}^n$ as the operator specified by the series

$$\exp(t\partial_\xi) = 1 + t\partial_\xi + \frac{t^2}{2}(\partial_\xi)^2 + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!}(\partial_\xi)^n.$$

Its action on smooth functions also is specified by the formal series

$$\exp(t\partial_\xi) f(x) = f(x) + t\partial_\xi f(x) + \frac{t^2}{2}(\partial_\xi)^2 f(x) + \cdots .$$

EXAMPLE. Let $\xi = \frac{d}{dx}$ be a constant vector field on the line $\mathbb{R}$. It generates the group of translations

$$\varphi_t f(x) = f(x+t) = f(x) + tf' + \frac{t^2}{2}f'' + \cdots ,$$

and its exponential function is

$$\exp\Big(t\frac{d}{dx}\Big) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\frac{d^n}{dx^n}.$$

The Taylor series is not always convergent, hence the exponential function specified by a formal series need not determine the action of a one-parameter group of diffeomorphisms. When the functions and the vector fields are specified by convergent Taylor series, the exponential mapping becomes well defined, as the following theorem shows.

**Theorem 8.11.** *If the vector field $\xi(x)$ is real-analytic, i.e., all its components $\xi^i$ are given by real-analytic functions of coordinates $x^1, \ldots, x^n$, then the action of the exponential function $\exp(t\partial_\xi)$ on real-analytic functions $f(x)$ coincides with the action (regular representation) of $\varphi_t$ for small $t$:*

$$\exp(t\partial_\xi)f(x) = \varphi_t f(x) = f(\varphi_t(x)).$$

**Proof.** If $\xi(x_0) = 0$, then $\exp(t\partial_\xi) = 1$ and $\varphi_t(x_0) = x_0$ at this point. Consider a point $x_0$ such that $\xi(x_0) \neq 0$. Then in a small neighborhood of this point the vector field is linearized in new coordinates $y = y(x)$:

$$\xi = \frac{\partial}{\partial y^1}.$$

Moreover, if the field $\xi$ is real-analytic, then the change of coordinates may be chosen real-analytic. In the new coordinates

$$\varphi_t f(y) = f(y + t\xi), \quad \exp(t\partial_\xi) = \sum \frac{t^n}{n!} \frac{\partial^n}{\partial (y^1)^n},$$

and the equality $\exp(t\partial_\xi)f(x) = \varphi_t \circ f(x)$ holds for small $t$ (i.e., in a domain where the coordinates $(y)$ are defined and the Taylor series for $f$ converges to $f$). Hence the theorem. $\square$

Thus we conclude that if we know the generators of a matrix group, then with the aid of the exponential mapping we can recover the (regular) action of the group on real-analytic functions:

$$\exp(t\partial_{T_A})f(x) = f(\exp(-tA)x).$$

This holds for infinite-dimensional groups of diffeomorphisms as well, which in many applications are specified by means of their generators.

**8.3.5. Invariant fields on Lie groups.** Let $G$ be a matrix Lie group realized by $n \times n$ matrices. Consider the imbedding $G \subset \mathrm{GL}(n)$ of this group into the linear space $\mathfrak{gl}(n) = \mathbb{R}^{n^2}$ formed by all matrices of order $n$.

For each matrix $X \in \mathfrak{gl}(n)$ we can construct the linear field $R_X$ taking the value

$$R_X(A) = XA$$

at a point $A \in \mathbb{R}^{n^2}$. Theorem 8.9 implies that

$$[R_X, R_Y] = -R_{[X,Y]}. \tag{8.14}$$

By Theorem 8.10 the integral curves of this field have the form

$$A(t) = \exp(tX)A_0, \quad A(0) = A_0.$$

The field $R_X$ is *right-invariant*. This means that the field $R_X$ is invariant relative to the the actions of the group $G$ by right multiplications on the space $\mathfrak{gl}(n) = \mathbb{R}^{n^2}$,

$$R_X(A)B = R_X(AB), \quad A \in \mathbb{R}^{n^2}, \quad B \in G.$$

In a similar way one shows that the vector field

$$L_X(A) = AX$$

is *left-invariant*, i.e., $BL_X(A) = L_X(BA)$. Its integral curves have the form $A_0 \exp(tX)$, and

$$[L_X, L_Y] = L_{[X,Y]}. \tag{8.15}$$

Let a matrix $X$ belong to the Lie algebra of the Lie group $G$, $X \in \mathfrak{g}$. Then, as was shown in 5.2.4 (Lemma 5.11), any matrix $e^{tX}$ belongs to the group $G$. The vector $R_X = XA$ is tangent to the curve $e^{tX}A$ at the point $t = 0$, and the vector $L_X = AX$ is tangent to $Ae^{tX}$ at $t = 0$. Hence if $A \in G$, then the vectors $R_X(A)$ and $L_X(A)$ are tangent to the submanifold $G$ at the point $A$. Thus we have proved the following.

**Theorem 8.12.** *Let $X \in \mathfrak{g}$. Then the vector fields $L_X$ and $R_X$ are tangent to the submanifold $G$, i.e., they specify vector fields on this submanifold.*

This implies that these vector fields are obtained by translations of Lie groups, which were introduced in 6.1.4,

$$l_{A*}X = L_X, \quad r_{A*}X = R_X, \quad A \in G, \quad X \in \mathfrak{g}.$$

Since the translations are isomorphisms of tangent spaces, we, obviously, have the following result.

**Corollary 8.6.** *Let $X_1, \dots, X_n$ be a basis in $\mathfrak{g}$. Then $L_{X_1}(A), \dots, L_{X_n}(A)$ and $R_{X_1}(A), \dots, R_{X_n}(A)$ are bases in the tangent space $T_AG$ to the submanifold $G$ at the point $A$. In particular, at each point of the group $G$ the left- (or right-) invariant vector fields form the entire tangent space to the group at this point.*

The formulas (8.14) and (8.15) imply the following result.

**Theorem 8.13.** *The left-invariant vector fields on the group $G$ with commutation operation form a Lie algebra isomorphic to the Lie algebra $\mathfrak{g}$ of this group.*

*The right-invariant vector fields on the group $G$ with "commutation" operation*

$$R_X, R_Y \to -[R_X, R_Y]$$

*also form a Lie algebra isomorphic to the Lie algebra $\mathfrak{g}$.*

**8.3.6. The Lie derivative.** Let $\xi$ be a smooth vector field on a manifold $M^n$, and let $\varphi_t$ be the one-parameter local group of diffeomorphisms formed by translations along the integral curves of the field $\xi$.

To each point $x \in M^n$ corresponds the differential of the mapping

$$\varphi_{t*}\colon T_x M^n \to T_{\varphi_t(x)} M^n,$$

which defines a linear mapping of tangent vectors by the formula

$$\varphi_{t*}(\xi)^i = \xi^j \frac{\partial \varphi_t^i}{\partial x^j}.$$

Since $\varphi_t$ is a diffeomorphism, this differential is an isomorphism of tangent spaces at the points $x$ and $\varphi_t(x)$. The linear mapping conjugate to it also is an isomorphism, but in this case, of cotangent spaces,

$$\varphi_t^*\colon T^*_{\varphi_t(x)} M^n \to T^*_x M^n.$$

We introduced this mapping in a more general setup in 8.2.2.

Define the "translation" of tensors along the integral curves from the point $\varphi_t(x)$ into the point $x$ as follows. If $T$ is a tensor of zero rank, i.e., a function, then it is natural to set

$$\varphi_t(T(x)) = T(\varphi_t(x)).$$

If we have a tensor

$$T = e_{(i)}^{(j)} = e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e^{j_1} \otimes \cdots \otimes e^{j_l}$$

of type $(k, l)$ at the point $\varphi_t(x)$, then its translation (in time $t$) along the integral curve is the tensor of the same type,

$$\varphi_t T = \varphi_{t*}^{-1}(e_{i_1}) \otimes \cdots \otimes \varphi_{t*}^{-1}(e_{i_k}) \otimes \varphi_t^*(e^{j_1}) \otimes \cdots \otimes \varphi_t^*(e^{j_l}),$$

at the point $x$. This mapping is extended by linearity to all tensors. According to this definition the translation commutes with the operation of tensor contraction.

As a result we obtain a smooth family of tensors $\varphi_t T$ at the point $x$.

The *Lie derivative* of the tensor field $T^{i_1\dots i_k}_{j_1\dots j_l}$ in the direction of the vector field $\xi$ is defined by the formula

$$L_\xi T^{i_1\dots i_k}_{j_1\dots j_l} = \frac{d}{dt} (\varphi_t T)^{i_1\dots i_k}_{j_1\dots j_l}\Big|_{t=0}.$$

It determines the rate of variation of $T$ under deformations of the space produced by the mappings $\varphi_t$. In applications (e.g., in continuum mechanics), the tensor field depends explicitly on time as well, $T = T(t, x)$. The expression

$$\frac{dT}{dt} = \frac{\partial T}{\partial t} + L_\xi T$$

is called the *total derivative* of the tensor $T$ in the direction of the field of velocities $\xi$.

The definition of the Lie derivative entails the following lemma.

**Lemma 8.3.** 1. *The Lie derivative does not affect the type of the tensor: if $T$ is a tensor of type $(k,l)$, then $L_\xi T$ is also a tensor of type $(k,l)$.*

2. *The Lie derivative commutes with contraction of tensors.*

3. *For the Lie derivative, the Leibniz formula holds:*

$$L_\xi(R\otimes S) = (L_\xi R)\otimes S + R\otimes (L_\xi S).$$

4. *If $\omega_1$ and $\omega_2$ are differential forms, then*

$$L_\xi(\omega_1\wedge\omega_2) = (L_\xi\omega_1)\wedge\omega_2 + \omega_1\wedge(L_\xi\omega_2).$$

**Proof.** Statements 1 to 3 follow directly from the definition. Statement 4 follows from the Leibniz formula, since the exterior product is by definition the alternating sum of tensor products. □

For small $t$ the points $x$ and $\varphi_t(x)$ lie in the same domain with coordinates $x^1,\dots,x^n$. Hence the tensors attached to different points will be written in the same coordinates.

According to (8.12) the mapping $\varphi_{t*}^{-1} = \varphi_{-t*}$ has the form

$$T^i \to \Big(\delta^i_j - t\,\frac{\partial\xi^i(\varphi_t(x))}{\partial x^j} + o(t)\Big)T^j = \Big(\delta^i_j - t\,\frac{\partial\xi^i(x)}{\partial x^j} + o(t)\Big)T^j.$$

Therefore, the mapping of covectors has the form

$$(\varphi_t^*T)_i = \Big(\delta^j_i + t\frac{\partial\xi^j(x)}{\partial x^i} + o(t)\Big)T_j.$$

Now we can easily obtain the formula for translation of an arbitrary tensor:

$$\begin{aligned}
(\varphi_t T)^{i_1\dots i_r}_{j_1\dots j_s} &= T^{k_1\dots k_r}_{l_1\dots l_s}\Big(\delta^{l_1}_{j_1} + t\,\frac{\partial\xi^{l_1}}{\partial x^{j_1}}\Big)\cdots\Big(\delta^{l_s}_{j_s} + t\,\frac{\partial\xi^{l_s}}{\partial x^{j_s}}\Big)\\
&\quad\times\Big(\delta^{i_1}_{k_1} - t\,\frac{\partial\xi^{i_1}}{\partial x^{k_1}}\Big)\cdots\Big(\delta^{i_r}_{k_r} - t\,\frac{\partial\xi^{i_r}}{\partial x^{k_r}}\Big)\\
&= T^{i_1\dots i_r}_{j_1\dots j_s} + t\Big(T^{i_1\dots i_r}_{m j_2\dots j_s}\frac{\partial\xi^m}{\partial x^{j_1}} + \dots + T^{i_1\dots i_r}_{j_1\dots j_{s-1}m}\frac{\partial\xi^m}{\partial x^{j_s}}\\
&\quad - T^{m i_2\dots i_r}_{j_1\dots j_s}\frac{\partial\xi^{i_1}}{\partial x^m} - \dots - T^{i_1\dots i_{r-1}m}_{j_1\dots j_s}\frac{\partial\xi^{i_r}}{\partial x^m}\Big) + o(t).
\end{aligned}$$

Here the tensor in the right-hand side is evaluated at the point $\varphi_t(x)$. Differentiating this equality with respect to $t$ at $t=0$, we obtain the formula

for the Lie derivative of the tensor field $T$:

$$(L_\xi T)^{i_1\dots i_r}_{j_1\dots j_s} = \xi^m \frac{\partial T^{i_1\dots i_r}_{j_1\dots j_s}}{\partial x^m} + T^{i_1\dots i_r}_{mj_2\dots j_s}\frac{\partial \xi^m}{\partial x^{j_1}} + \dots + T^{i_1\dots i_r}_{j_1\dots j_{s-1}m}\frac{\partial \xi^m}{\partial x^{j_s}} - T^{mi_2\dots i_r}_{j_1\dots j_s}\frac{\partial \xi^{i_1}}{\partial x^m} - \dots - T^{i_1\dots i_{r-1}m}_{j_1\dots j_s}\frac{\partial \xi^{i_r}}{\partial x^m}. \tag{8.16}$$

EXAMPLES. 1. FUNCTIONS (TENSORS OF RANK 0). The Lie derivative for functions coincides with the directional derivative in the direction of the vector field:

$$L_\xi f = \xi^i \frac{\partial f}{\partial x^i} = \partial_\xi f.$$

The function $f$ is constant along the flow $\varphi_t$ if and only if $L_\xi f = 0$. In this case $f$ is called the *integral of the field* $\xi$ or the integral of the equation $\dot{x} = \xi(x)$. If the equation $\dot{x} = \xi(x)$ has the first integral $f$, then its order can by reduced by one in the following sense. The integral curves lie on the level surfaces $f(x^1, \dots, x^n) = \text{const}$, and the field $\xi$ is, obviously, tangent to these surfaces. The level surfaces are of dimension $(n-1)$, and the restriction of the equation $\dot{x} = \xi(x)$ to them yields an equation involving fewer variables $x^i$.

2. VECTORS. The equality (8.16) implies that

$$L_\xi \eta^i = \xi^j \frac{\partial \eta^i}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j} = [\xi^i, \eta^i],$$

i.e., $L_\xi \eta$ coincides with the commutator of the vector fields $\xi$ and $\eta$.

3. COVECTORS. It follows from (8.16) that

$$(L_\xi \eta)_j = \xi^i \frac{\partial \eta_j}{\partial x^i} + \eta_i \frac{\partial \xi^i}{\partial x^j}.$$

**Corollary 8.7.** *For functions, the differential commutes with the Lie derivative,* $L_\xi(df) = d(L_\xi f)$.

**Proof.** Let $\eta_i = \frac{\partial f}{\partial x^i}$ be the gradient of the function $f$. Then

$$(L_\xi \eta)_j = \xi^i \frac{\partial^2 f}{\partial x^i \partial x^j} + \frac{\partial \xi^i}{\partial x^j}\frac{\partial f}{\partial x^i} = \frac{\partial}{\partial x^j}\partial_\xi f.$$

Hence the corollary. □

4. TENSORS OF TYPE $(0, k)$ AND, IN PARTICULAR, DIFFERENTIAL FORMS. The Lie derivative for them is written as

$$L_\xi T = \frac{d}{dt}\varphi_t^* T\Big|_{t=0}.$$

For example, for bilinear forms we have

$$(L_\xi g)_{ij} = \xi^k \frac{\partial g_{ij}}{\partial x^k} + \frac{\partial \xi^k}{\partial x^i} g_{kj} + \frac{\partial \xi^k}{\partial x^j} g_{ik} = u_{ij}. \tag{8.17}$$

For a metric $g_{ij}$ the tensor $L_\xi g$ is called the *small-deformation tensor*. It describes the variation of the metric under the deformation of the space specified by the flow $\varphi_t$. For the Euclidean metric $g_{ij} = \delta_{ij}$ (in which case the superscripts and subscripts are undistinguishable) this tensor takes the standard form for elasticity theory,

$$u_{ij} = \frac{\partial \xi^i}{\partial x^j} + \frac{\partial \xi^j}{\partial x^i}.$$

5. THE VOLUME ELEMENT. Consider the volume element $\sqrt{|g|}\varepsilon_{i_1\dots i_n}$, where $g = \det g_{ik}$, which transforms as a tensor only under coordinate changes with positive Jacobian. Formula (8.16) determines its Lie derivative, which equals

$$\begin{aligned} L_\xi \sqrt{|g|}\varepsilon_{i_1\dots i_n} &= \xi^k \frac{\partial\sqrt{|g|}}{\partial x^k}\varepsilon_{i_1\dots i_n} \\ &\quad + \sqrt{|g|}\left(\varepsilon_{k i_2\dots i_n}\frac{\partial \xi^k}{\partial x^{i_1}} + \dots + \varepsilon_{i_1\dots i_{n-1}k}\frac{\partial \xi^k}{\partial x^{i_n}}\right) \\ &= \xi^k \frac{\partial\sqrt{|g|}}{\partial x^k}\varepsilon_{i_1\dots i_n} + \sqrt{|g|}\,\frac{\partial \xi^i}{\partial x^i}\varepsilon_{i_1\dots i_n} \\ &= \sqrt{|g|}\,\varepsilon_{i_1\dots i_n}\left(\xi^k \frac{\partial \log\sqrt{|g|}}{\partial x^k} + \frac{\partial \xi^i}{\partial x^i}\right) \\ &= \frac{1}{2}\sqrt{|g|}\,\varepsilon_{i_1\dots i_n}\left(\xi^k g^{ij}\frac{\partial g_{ij}}{\partial x^k} + 2\frac{\partial \xi^i}{\partial x^i}\right). \end{aligned}$$

Note that the expression in parentheses is the trace of the deformation tensor $\mathrm{Tr}(u_{ij}) = g^{ij}u_{ij}$ specified by (8.17). As a result, we obtain

$$L_\xi(\sqrt{|g|}\varepsilon_{i_1\dots i_n}) = \frac{1}{2}\,g^{ij}u_{ij}\sqrt{|g|}\varepsilon_{i_1\dots i_n}.$$

For the Euclidean metric the right-hand side simplifies:

$$L_\xi(\sqrt{|g|}\varepsilon_{i_1\dots i_n}) = \frac{\partial \xi^i}{\partial x^i}\,\varepsilon_{i_1\dots i_n}.$$

**8.3.7. Central extensions of Lie algebras.** Important for physical applications are not only Lie algebras of vector fields, but also their central extensions. The most widely known among them is the Virasoro algebra.

A Lie algebra $L^+$ is said to be a *central extension* of an algebra $L$ if it is obtained from $L$ by adding the space $V$ generated by elements $t_1, \dots, t_n$ with commutation relations

$$[\xi,\eta]_+ = [\xi,\eta] + f_1(\xi,\eta)t_1 + \dots + f_n(\xi,\eta)t_n, \quad [L^+, t_i]_+ = 0,\ i = 1,\dots,n,$$

where $\xi, \eta \in L$, $[\cdot,\cdot]$ is the commutator in the algebra $L$, and $[\cdot,\cdot]_+$ is the commutator in the algebra $L^+$. Thus we see that the subspace $V$ consists of central elements of $L^+$ (i.e., commuting with all others).

Obviously, any (finite-dimensional) central extension can be obtained by successive one-dimensional central extensions (with $\dim V = 1$), each specified by a function

$$f\colon L \times L \to F,$$

where

$$[\xi, \eta]_+ = [\xi, \eta] + f(\xi, \eta)t$$

and $F$ is the field of coefficients of the Lie algebra. The function $f$ must be bilinear and skew-symmetric to provide for anticommutativity $[\xi, \eta]_+ = -[\eta, \xi]_+$. It can be easily verified that the following condition is necessary and sufficient for the Jacobi identity to hold:

$$f([\xi, \eta], \tau) + f([\eta, \tau], \xi) + f([\tau, \xi], \eta) = 0, \quad \xi, \eta, \tau \in L. \tag{8.18}$$

This condition is satisfied, e.g., for functions of the form

$$f(\xi, \eta) = g([\xi, \eta]), \tag{8.19}$$

where $g$ is any linear function on $L$.

A bilinear skew-symmetric function $f(\xi, \eta)$ satisfying condition (8.18) is called a (two-dimensional) *cocycle* on the Lie algebra. A cocycle of the form (8.19) is called a *coboundary*.

Obviously, the cocycles form an additive group $Z(L)$, and the coboundaries comprise its subgroup $B(L) \subset Z(L)$.

Each coboundary specifies a trivial central extension of the Lie algebra. Indeed, if $L^+$ is a central extension determined by a coboundary $g$, then there exists an isomorphism

$$L^+ \to L \oplus Ft: \ \xi \to \xi + g(\xi)t, \ t \to t, \ \xi \in L.$$

It can be shown in a similar way that if two cocycles $f_1$ and $f_2$ differ by a coboundary $g = f_1 - f_2$, then they specify isomorphic extensions, and that a one-dimensional central extension is isomorphic to a trivial extension if and only if the cocycle that specifies this extension is a coboundary. As a conclusion, we state the following theorem.

**Theorem 8.14.** *One-dimensional central extensions of a Lie algebra $L$ are in one-to-one correspondence (up to isomorphism) with the commutative group $H^2(L)$, which is the quotient group of the group of cocycles by the subgroup of coboundaries,*

$$H^2(L) = Z(L)/B(L).$$

The group $H^2(L)$ is called the (*two-dimensional*) *cohomology group* of the Lie algebra $L$.

Smooth vector fields on the circle have the form

$$\xi = \xi(x)\,\frac{d}{dx}, \quad \xi(x + 2\pi) = \xi(x).$$

The function $\xi(x)$ can be expanded in a Fourier series in $\sin nx$ and $\cos nx$, $n = 0, 1, \ldots$, hence the vector fields

$$\frac{d}{dx}, \quad \cos nx \, \frac{d}{dx}, \quad \sin nx \, \frac{d}{dx}, \quad n = 1, 2, \ldots,$$

compose a formal basis in the Lie algebra $V(S^1)$ of smooth vector fields on the circle.

In order to simplify the commutation relations, it is expedient to consider the complexification $V(S^1) \otimes \mathbb{C}$ of the algebra $V(S^1)$, which is formally generated by the basis elements

$$e_n = e^{inx} \frac{d}{dx}, \quad n \in \mathbb{Z},$$

with commutation relations

$$[e_m, e_n] = i(n - m)e_{m+n}, \quad m, n \in \mathbb{Z}.$$

The Lie algebra $V(S^1)$ is generated over the field of real numbers by the fields $e_n + e_{-n}$ and $i(e_n - e_{-n})$, $n \geq 0$.

**Lemma 8.4.** *The bilinear skew-symmetric function*

$$f(\xi, \eta) = \int_0^{2\pi} \xi''' \eta \, dx \tag{8.20}$$

*is a cocycle on the Lie algebra $V(S^1)$ and on its complexification $V(S^1) \otimes \mathbb{C}$.*

Here the prime denotes the derivative with respect to $x$.

**Proof.** The function $f(\xi, \eta)$ takes the following values on the elements of the basis $\{e_n\}$:

$$f(e_n, e_m) = -2\pi i n^3 \delta_{n,-m}.$$

We have

$$\begin{aligned} &f([e_k, e_l], e_m) + f([e_l, e_m], e_k) + f([e_m, e_k], e_l) \\ &\quad = 2\pi \delta_{k+l+m,0}((l - k)(k + l)^3 + (m - l)(l + m)^3 + (k - m)(k + m)^3) \\ &\quad = 0. \end{aligned}$$

Hence (8.20) specifies a cocycle on the algebra $V(S^1) \otimes \mathbb{C}$. Since it is real-valued on the subalgebra $V(S^1)$, it specifies a cocycle on the algebra $V(S^1)$ as well. □

The cocycle (8.20) is referred to as the *Gelfand–Fuchs cocycle.* The following theorem due to I. Gelfand and D. Fuchs is stated without proof.

**Theorem 8.15.** *The cohomology group $H^2(V(S^1))$ is one-dimensional, $H^2(V(S^1)) = \mathbb{R}$, and its generator is the cocycle* (8.20).

For the Lie algebra of vector fields on the circle, the one-dimensional central extension $V_c$ of the form

$$\left[\xi \frac{d}{dx}, \eta \frac{d}{dx}\right]_+ = (\xi\eta' - \xi'\eta)\frac{d}{dx} + 2\pi c\left(\int_0^{2\pi} \xi'''\eta\,dx\right)t, \quad \left[\xi \frac{d}{dx}, t\right] = 0,$$

is called the *Virasoro algebra.* The constant $c$ is called the *central charge* of the extension $V_c$.

## Exercises to Chapter 8

1. Find the action of the operator $*$ on skew-symmetric tensors in the 4-dimensional Minkowski space.

2. Prove that the action (8.4) of the Lorentz group on the space of skew-symmetric tensors of rank 2 in the Minkowski space specifies a Lie group isomorphism, $\mathrm{SO}(1,3) = \mathrm{O}(3,\mathbb{C})$.

3. Show that the electromagnetic field in the Minkowski space is determined by a pair of tensors consisting of a vector in $\mathbb{R}^3$ and a tensor that is skew-symmetric relative to changes of spacial coordinates (which do not affect time).

4. Prove the following formula:

$$F^*(\omega_1 \wedge \omega_2) = F^*(\omega_1) \wedge F^*(\omega_2).$$

5. Prove that for diffeomorphisms $F\colon U \to V$ the mappings $F_*$ take commutators of vector fields again into commutators,

$$F_*[\xi, \eta] = [F_*\xi, F_*\eta].$$

6. Prove that one-parameter groups commute if and only if the vector fields $\xi$ and $\eta$ generating them commute,

$$[\xi, \eta] = 0.$$

7. Prove that if linearly independent vector fields $X_1, \dots, X_n$ in a domain in $\mathbb{R}^n$ commute, i.e., $[X_i, X_j] = 0$, then there are coordinates $x^1, \dots, x^n$ in this domain such that each field $X^i$ is everywhere tangent to the $i$th coordinate axis, $\partial_{X_i}(x^j) = \delta_i^j$.

8. Compute the operator $\exp[(ax+b)\frac{d}{dx}]$.

9. Write down the generators of the group of affine transformations of $\mathbb{R}^n$ and the group of motions of the space $\mathbb{R}^{1,n}$.

10. Prove that the vector fields $L_X$, $L_Y$, $L_Z$ in $\mathbb{R}^3$, which generate the group $\mathrm{SO}(3,\mathbb{R})$, are tangent to any sphere with center at the origin and find the form of the corresponding differential operators of first order on the unit sphere in spherical coordinates.

11. Prove that left-invariant vector fields $L_X$ on a Lie group commute with right-invariant vector fields $R_Y$:

$$[L_X, R_Y] \equiv 0.$$

12. Let $L_\xi$ and $L_\eta$ be the Lie derivatives in the direction of vector fields $\xi$ and $\eta$. Prove the formula

$$L_\xi L_\eta - L_\eta L_\xi = L_{[\xi,\eta]}.$$

13. A vector field $\xi$ on a Riemannian manifold with metric $g_{ik}$ is called a *Killing field* if the Lie derivative of the metric in the direction of this field vanishes,

$$L_\xi g_{ik} = 0.$$

Prove that:

a) If $\varphi_t$ is a one-parameter group of isometries acting on the manifold, then the field $\xi(x) = \dot{\varphi}_t(x)\big|_{t=0}$ tangent to it is a Killing field.

b) A linear combination of two Killing fields with constant coefficients is a Killing field.

c) The commutator of two Killing fields is a Killing field.

Chapter 9

# Analysis of Differential Forms

## 9.1. Differential forms

**9.1.1. Skew-symmetric tensors and their differentiation.** Physical laws (at least, most of them) are expressed as differential relations between physical quantities representable as tensor fields. This is due to the fact that physical laws are independent of the choice of coordinates and hence are of tensor nature. Differential operations involved in them must also be independent of the choice of a particular coordinate system, hence any such operation applied to a tensor must result again in a tensor.

For example, let $T(x^1, \dots, x^n, \alpha)$ be a tensor field depending on a parameter $\alpha$. Then the derivative $\partial T/\partial \alpha$ with respect to this parameter is a tensor of the same type as $T$. This is because differentiation is performed at each particular point and is unrelated to the geometry of the space. In classical mechanics, the role of parameter $\alpha$ is played by time, $\alpha = t$.

Taking the gradient of a smooth function,

$$f(x^1, \dots, x^n) \to \operatorname{grad} f = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}\right),$$

carries the function (scalar) $f$ into the covector $\operatorname{grad} f$, which transforms under coordinate changes $z = z(x)$ by the rule

$$\frac{\partial f}{\partial z^i} = \frac{\partial f}{\partial x^j}\frac{\partial x^j}{\partial z^i}.$$

This operation is not connected with the metric (geometry) of the space either.

In physics of the 19th century two more operations arose on vector fields in three-dimensional space with Euclidean coordinates $x^1, x^2, x^3$ (we will not distinguish between superscripts and subscripts of coordinates in Euclidean space). With a vector field $T = (T_1, T_2, T_3)$ one associates its *curl*

$$\operatorname{curl} T = \Big(\frac{\partial T_3}{\partial x^2} - \frac{\partial T_2}{\partial x^3}, \frac{\partial T_1}{\partial x^3} - \frac{\partial T_3}{\partial x^1}, \frac{\partial T_2}{\partial x^1} - \frac{\partial T_1}{\partial x^2}\Big)$$

and *divergence*

$$\operatorname{div} T = \frac{\partial T_1}{\partial x^1} + \frac{\partial T_2}{\partial x^2} + \frac{\partial T_3}{\partial x^3}.$$

Under transitions to other Euclidean coordinates, the curl transforms as a vector field, while divergence behaves as a scalar. Since we do not distinguish between vectors and covectors in Euclidean space, we may assume these operations to be defined on covectors as well. Note that they fulfill the following identities:

$$\operatorname{curl} \operatorname{grad} f = 0, \quad \operatorname{div} \operatorname{curl} T = 0.$$

These operations admit the extension to general skew-symmetric tensors of type $(0, k)$ as follows.

Let $T_{j_1 \dots j_k}$ be a skew-symmetric tensor of type $(0, k)$ in $n$-dimensional space with coordinates $x^1, \dots, x^n$, $j_m = 1, \dots, n$. The *gradient* of the tensor $T$ is the skew-symmetric tensor of type $(0, k+1)$ with components

$$(dT)_{j_1 \dots j_{k+1}} = \sum_{m=1}^{k+1} (-1)^{m+1} \frac{\partial T_{j_1 \dots \widehat{j_m} \dots j_{k+1}}}{\partial x^{j_m}}$$

(as usual, the "hat" over the subscript $j_m$ means that this subscript is skipped).

First of all, we explain why this operation generalizes the above examples.

1) If $T$ is a scalar (function), then $(dT)_i = \frac{\partial T}{\partial x^i}$ and we see that $dT = \operatorname{grad} T$.

2) To each covector field $T$ in $\mathbb{R}^3$ corresponds the vector field $dT$ of type $(0, 2)$. In three-dimensional Euclidean space, the *curl* of the field $T$ has the form

$$\operatorname{curl} T = * \, dT,$$

where $*$ is the Hodge star-operator.

3) For a tensor $T$ of type $(0, 2)$ in three-dimensional space we have

$$(dS)_{123} = \frac{\partial S_{12}}{\partial x^3} - \frac{\partial S_{13}}{\partial x^2} + \frac{\partial S_{23}}{\partial x^1},$$

which implies that in Euclidean space $\mathbb{R}^3$ the *divergence* of the field $T$ is

$$\operatorname{div} T = *^{-1} d(* T).$$

We verify that the notion of gradient of a tensor is well defined.

**Theorem 9.1.** *The gradient $dT$ of a skew-symmetric tensor of type $(0,k)$ is a skew-symmetric tensor of type $(0,k+1)$.*

**Proof.** For $k=0$ this theorem was proved in 7.1.1.

To avoid cumbersome calculations, we restrict ourselves to the case of $k=1$. An analysis of this proof makes it clear how to proceed in the general case.

Consider a change of coordinates

$$x^i = x^i(z^1,\dots,z^n), \quad i=1,\dots,n.$$

Let $T_{i_1\dots i_k}$ and $\tilde{T}_{j_1\dots j_k}$ be the components of the tensor in coordinates $(x)$ and $(z)$, respectively. By definition,

$$\tilde{T}_{j_1\dots j_k} = T_{i_1\dots i_k}\frac{\partial x^{i_1}}{\partial z^{j_1}}\cdots\frac{\partial x^{i_k}}{\partial z^{j_k}}. \tag{9.1}$$

Consider the gradients of the tensor in these coordinate systems:

$$(d\tilde{T})_{j_1\dots j_{k+1}} = \sum_m (-1)^{m+1}\frac{\partial \tilde{T}_{j_1\dots\widehat{j_m}\dots j_{k+1}}}{\partial z^{j_m}}, \tag{9.2}$$

$$(dT)_{i_1\dots i_{k+1}} = \sum_p (-1)^{p+1}\frac{\partial T_{i_1\dots\widehat{i_p}\dots i_{k+1}}}{\partial x^{i_p}}. \tag{9.3}$$

To prove the theorem, we must substitute formulas (9.2) and (9.3) into formula (9.1) of tensor transformation, in order to see that the gradient $(d\tilde{T})_{(j)}$ is expressed in terms of $(dT)_{(i)}$ by the tensor rule. Let $k=1$ and let $T_i$ be a covector. Then

$$(dT)_{ij} = \frac{\partial T_j}{\partial x^i} - \frac{\partial T_i}{\partial x^j}, \quad \tilde{T}_j = T_i\frac{\partial x^i}{\partial z^j}.$$

We have

$$\begin{aligned}
(d\tilde{T})_{lk} &= \frac{\partial \tilde{T}_k}{\partial z^l} - \frac{\partial \tilde{T}_l}{\partial z^k} = \frac{\partial}{\partial z^l}\Big(T_i\frac{\partial x^i}{\partial z^k}\Big) - \frac{\partial}{\partial z^k}\Big(T_j\frac{\partial x^j}{\partial z^l}\Big)\\
&= \frac{\partial T_i}{\partial z^l}\frac{\partial x^i}{\partial z^k} + T_i\frac{\partial^2 x^i}{\partial z^l\partial z^k} - \frac{\partial T_j}{\partial z^k}\frac{\partial x^j}{\partial z^l} - T_j\frac{\partial^2 x^j}{\partial z^k\partial z^l}\\
&= \Big(\frac{\partial T_i}{\partial x^j}\frac{\partial x^j}{\partial z^l}\Big)\frac{\partial x^i}{\partial z^k} - \Big(\frac{\partial T_j}{\partial x^i}\frac{\partial x^i}{\partial z^k}\Big)\frac{\partial x^j}{\partial z^l}\\
&= \Big(\frac{\partial T_i}{\partial x^j} - \frac{\partial T_j}{\partial x^i}\Big)\frac{\partial x^i}{\partial z^k}\frac{\partial x^j}{\partial z^l} = (dT)_{ji}\frac{\partial x^i}{\partial z^k}\frac{\partial x^j}{\partial z^l},
\end{aligned}$$

which proves the theorem for $k=1$. □

**9.1.2. Exterior differential.** Let $T_{i_1 \dots i_k}$ be a skew-symmetric tensor of type $(0,k)$. It was shown in 7.2.3 that one can associate with it the differential form

$$\omega = \sum_{i_1 < \dots < i_k} T_{i_1 \dots i_k}\, dx^{i_1} \wedge \dots \wedge dx^{i_k}.$$

The exterior differential of the form $\omega$ is the form $d\omega$ of degree $k+1$ defined by the formula

$$d\omega = \sum_{i_0,\ i_1 < \dots < i_k} \frac{\partial T_{i_1 \dots i_k}}{\partial x^{i_0}}\, dx^{i_0} \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k}. \tag{9.4}$$

If $\omega = f$ is a scalar, then $d\omega = \frac{\partial f}{\partial x^i}\, dx^i$ is the differential of the function $f$.

**Theorem 9.2.** *If* $\omega = \sum_{i_1 < \dots < i_k} T_{i_1 \dots i_k}\, dx^{i_1} \wedge \dots \wedge dx^{i_k}$, *then*

$$d\omega = \sum_{j_1 < \dots < j_{k+1}} (dT)_{j_1 \dots j_{k+1}}\, dx^{j_1} \wedge \dots \wedge dx^{j_{k+1}}.$$

**Proof.** By the definition of $dT$ we have

$$\begin{aligned} &\sum_{j_1 < \dots < j_{k+1}} (dT)_{j_1 \dots j_{k+1}}\, dx^{j_1} \wedge \dots \wedge dx^{j_{k+1}} \\ &\qquad = \sum_{j_1 < \dots < j_{k+1}} \sum_m (-1)^{m+1} \frac{\partial T_{j_1 \dots \widehat{j_m} \dots j_{k+1}}}{\partial x^{j_m}}\, dx^{j_1} \wedge \dots \wedge dx^{j_{k+1}}. \end{aligned}$$

Since $dx^{j_1} \wedge \dots \wedge dx^{j_{k+1}} = (-1)^{m+1} dx^{j_m} \wedge dx^{j_1} \wedge \dots \wedge \widehat{dx^{j_m}} \wedge \dots \wedge dx^{j_{k+1}}$, by setting $i_0 = j_m$, $i_1 = j_1, \dots, i_k = j_{k+1}$ in the $m$th term the right-hand side is reduced to the form (9.4). □

The identities $\operatorname{curl} \operatorname{grad} f = 0$ and $\operatorname{div} \operatorname{curl} T = 0$ are generalized as follows.

**Theorem 9.3.** *The squared exterior differential of a form is identically equal to zero,*

$$d^2\omega = d(d\omega) = 0.$$

**Proof.** For a differential form

$$\omega = \sum_{i_1 < \dots < i_k} T_{i_1 \dots i_k}\, dx^{i_1} \wedge \dots \wedge dx^{i_k}$$

we have

$$d^2\omega = d(d\omega) = \sum_{p,q,\ i_1 < \dots < i_k} \frac{\partial^2 T_{i_1 \dots i_k}}{\partial x^q \partial x^p}\, dx^q \wedge dx^p \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k}.$$

The expression $\frac{\partial^2 T_{i_1\dots i_k}}{\partial x^q \partial x^p}$ is symmetric in $q$ and $p$, whereas $dx^q \wedge dx^p$ is skew-symmetric, $dx^p \wedge dx^q = -dx^q \wedge dx^p$. This implies that their contraction vanishes. □

DEFINITION. A form $\omega$ is said to be *closed* if $d\omega = 0$, and it is said to be *exact* if $\omega = d\eta$ for some form $\eta$.

Theorem 9.3 asserts that all exact forms are closed.

The following theorem shows how the exterior differential acts on the exterior product of differential forms.

**Theorem 9.4.** *If $\omega_1$ and $\omega_2$ are differential forms of degree $p$ and $q$, respectively, then*

$$d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^p \omega_1 \wedge d\omega_2.$$

**Proof.** Let

$$\omega_1 = f\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}, \quad \omega_2 = g\, dx^{j_1} \wedge \cdots \wedge dx^{j_q}.$$

We have

$$\omega_1 \wedge \omega_2 = fg\, dx^{i_1} \wedge \cdots \wedge dx^{i_p} \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_q},$$

$$\begin{aligned}
d(\omega_1 \wedge \omega_2) &= \frac{\partial f}{\partial x^k} g\, dx^k \wedge dx^{i_1} \wedge \cdots \wedge dx^{j_q} + f \frac{\partial g}{\partial x^k}\, dx^k \wedge dx^{i_1} \wedge \cdots \wedge dx^{j_q} \\
&= \Big(\frac{\partial f}{\partial x^k}\, dx^k \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}\Big) \wedge (g\, dx^{j_1} \wedge \cdots \wedge dx^{j_q}) \\
&\quad + (-1)^p (f\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}) \wedge \Big(\frac{\partial g}{\partial x^k}\, dx^k \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_q}\Big) \\
&= d\omega_1 \wedge \omega_2 + (-1)^p \omega_1 \wedge d\omega_2,
\end{aligned}$$

since

$$\begin{aligned}
&dx^k \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p} \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_q} \\
&\qquad = (-1)^p\, dx^{i_1} \wedge \cdots \wedge dx^{i_p} \wedge dx^k \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_q}.
\end{aligned}$$

The proof is extended to general differential forms by linearity of $d$. □

We will give another expression for the differential of a form, which is useful in applications.

**Theorem 9.5.** *Let $\omega$ be a differential $k$-form, and $X_1, \dots, X_{k+1}$ smooth vector fields. Then*

$$\begin{aligned}
&d\omega(X_1, \dots, X_{k+1}) \\
&\quad = \sum_i (-1)^{i-1} \partial_{X_i} \omega(X_1, \dots, \widehat{X_i}, \dots, X_{k+1}) \\
&\qquad + \sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \dots, \widehat{X_i}, \dots, \widehat{X_j}, \dots, X_{k+1}).
\end{aligned} \tag{9.5}$$

Here the expression $\partial_{X_i}\omega(X_1,\dots,\widehat{X_i},\dots,X_{k+1})$ means the directional derivative of the smooth function $\omega(X_1,\dots,\widehat{X_i},\dots,X_{k+1})$ in the direction of the field $X_i$.

**Proof.** Let $T_i\,dx^i = \omega$; then

$$d\omega = \sum_{i<j}\Big(\frac{\partial T_j}{\partial x^i} - \frac{\partial T_i}{\partial x^j}\Big)\,dx^i\wedge dx^j$$

and the form $d\omega$ on the vector fields $X$ and $Y$ is equal to

$$(9.6)\quad d\omega(X,Y) = \sum_{i<j}\Big(\frac{\partial T_j}{\partial x^i} - \frac{\partial T_i}{\partial x^j}\Big)(X^iY^j - X^jY^i) = X^iY^j\Big(\frac{\partial T_j}{\partial x^i} - \frac{\partial T_i}{\partial x^j}\Big),$$

where the summation over repeated indices $i$ and $j$ is assumed in the right-hand side. For $k=1$ formula (9.5) becomes

$$\begin{aligned} d\omega(X,Y) &= \partial_X\omega(Y) - \partial_Y\omega(X) - \omega([X,Y]) \\ &= \partial_X(T_iY^i) - \partial_Y(T_iX^i) - T_i\Big(X^k\frac{\partial Y^i}{\partial x^k} - Y^k\frac{\partial X^i}{\partial x^k}\Big) \qquad (9.7)\\ &= X^iY^j\Big(\frac{\partial T_j}{\partial x^i} - \frac{\partial T_i}{\partial x^j}\Big). \end{aligned}$$

We see that the right-hand sides of (9.6) and (9.7) coincide. This proves the theorem for $k=1$. The general case is left to the reader. □

In the presence of a metric we can define one more operation on forms, which lowers the degree of the form by one, in contrast to the exterior differential, which increases it by one. This is the divergence $\delta$ of a skew-symmetric tensor. Its action on $k$-forms on an $n$-dimensional manifold is defined by the formula

$$\delta = (-1)^k *^{-1} d * = (-1)^{nk+n+1} * d *.$$

Theorem 9.3 implies that

$$\delta^2 = (*^{-1}d*)(*^{-1}d*) = *^{-1}d^2* = 0.$$

Example. We derive once more the formulas for curl and divergence in $\mathbb{R}^3$ with Euclidean coordinates $x$. Let $\omega = T_1\,dx^1 + T_2\,dx^2 + T_3\,dx^3$. Then

$$d\omega = \Big(\frac{\partial T_2}{\partial x^1}-\frac{\partial T_1}{\partial x^2}\Big)\,dx^1\wedge dx^2 + \Big(\frac{\partial T_3}{\partial x^2}-\frac{\partial T_2}{\partial x^3}\Big)\,dx^2\wedge dx^3 + \Big(\frac{\partial T_1}{\partial x^3}-\frac{\partial T_3}{\partial x^1}\Big)\,dx^3\wedge dx^1$$

is a 2-form, and

$$*\,d\omega = \Big(\frac{\partial T_3}{\partial x^2} - \frac{\partial T_2}{\partial x^3}\Big)\,dx^1 + \Big(\frac{\partial T_1}{\partial x^3} - \frac{\partial T_3}{\partial x^1}\Big)\,dx^2 + \Big(\frac{\partial T_2}{\partial x^1} - \frac{\partial T_1}{\partial x^2}\Big)\,dx^3.$$

On identifying vectors and covectors we obtain

$$\operatorname{curl} T = *\,dT.$$

Now we write down the form $\delta\omega = -*\,d*\omega$. This is a form of degree zero:

$$\begin{aligned}\omega &= T_1\,dx^1 + T_2\,dx^2 + T_3\,dx^3\\ &\xrightarrow{*} T_1\,dx^2\wedge dx^3 + T_2\,dx^3\wedge dx^1 + T_3\,dx^1\wedge dx^2\\ &\xrightarrow{d} \Big(\frac{\partial T_1}{\partial x^1}+\frac{\partial T_2}{\partial x^2}+\frac{\partial T_3}{\partial x^3}\Big)\,dx^1\wedge dx^2\wedge dx^3 \xrightarrow{-*} -\Big(\frac{\partial T_1}{\partial x^1}+\frac{\partial T_2}{\partial x^2}+\frac{\partial T_3}{\partial x^3}\Big).\end{aligned}$$

Therefore

$$\delta T = -*\,d*T = -\operatorname{div}T = -\Big(\frac{\partial T_1}{\partial x^1}+\frac{\partial T_2}{\partial x^2}+\frac{\partial T_3}{\partial x^3}\Big).$$

In non-Euclidean coordinates, divergence is determined by a different formula (see (10.31) below).

**9.1.3. Maxwell equations.** As an example of a physical law involving differential relations of curl and divergence type we present the Maxwell equations for electromagnetic field.

Consider the four-dimensional Minkowski space (space–time) with coordinates $x^0 = ct$, $x^1$, $x^2$, $x^3$ (where $c$ is the velocity of light) and pseudo-Euclidean metric

$$dl^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2.$$

The electromagnetic field is a skew-symmetric tensor $F_{ij}$ of rank 2, where $i,j = 0,1,2,3$ (see 8.1.1). It is determined by the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ according to the formula

$$(F_{ij}) = \begin{pmatrix} 0 & E_1 & E_2 & E_3\\ -E_1 & 0 & -H_3 & H_2\\ -E_2 & H_3 & 0 & -H_1\\ -E_3 & -H_2 & H_1 & 0\end{pmatrix}.$$

The tensor $F_{ij}$ of electromagnetic field satisfies the *Maxwell equations*, which fall into two pairs of equations.

The first pair is

$$(dF)_{ijk} = \frac{\partial F_{jk}}{\partial x^i} + \frac{\partial F_{ki}}{\partial x^j} + \frac{\partial F_{ij}}{\partial x^k} = 0. \tag{9.8}$$

In terms of the fields $\mathbf{E} = (F_{01}, F_{02}, F_{03})$ and $\mathbf{H} = (F_{32}, F_{13}, F_{21})$ equation (9.8) with $(i,j,k) = (1,2,3)$ becomes

$$\operatorname{div}\mathbf{H} = \frac{\partial F_{12}}{\partial x^3} + \frac{\partial F_{23}}{\partial x^1} + \frac{\partial F_{31}}{\partial x^2} = 0, \tag{9.9}$$

while the other equations in (9.8) mean that

$$\operatorname{curl}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}. \tag{9.10}$$

Thus we see that the system (9.8), or, equivalently,

$$dF = 0,$$

can be rewritten as a system consisting of the scalar equation (9.9) and the vector equation (9.10).

In contrast to the first pair, the second pair of Maxwell's equations is related to the presence of the pseudo-Euclidean metric and has the form

$$\delta F = *d*F = \frac{-4\pi}{c} j_{(4)}, \tag{9.11}$$

where $j_{(4)}$ is the four-dimensional (co)vector of current, $j_{(4)} = (\rho c, -\rho v^1, -\rho v^2, -\rho v^3) = (\rho, -\mathbf{j})$, $\rho$ is the density of the electric charge in the three-dimensional space, and $v = (v^1, v^2, v^3)$ is the usual velocity of charges in the three-dimensional space.

Using formula (8.3) for the operator $*$ in pseudo-Euclidean coordinates, equation (9.11) can be rewritten as a pair of equations: the scalar equation

$$\operatorname{div} \mathbf{E} = 4\pi\rho$$

and the vector equation

$$\operatorname{curl} \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{j}.$$

## 9.2. Integration of differential forms

**9.2.1. Definition of the integral.** In the classical calculus, for each continuous function $f(x^1, \ldots, x^n)$, we have its integral over a bounded domain $U$ of Euclidean space $\mathbb{R}^n$,

$$\int \cdots \int_U f(x)\, dx^1 \cdots dx^n = \int \cdots \int_U f(x)\, dx^1 \wedge \cdots \wedge dx^n.$$

If we are given a change of variables

$$y = y(x), \quad y \in V,$$

with positive Jacobian

$$J = \det\Big(\frac{\partial y^i}{\partial x^j}\Big) > 0,$$

then the following formula for the change of variables in the integral holds:

$$\int \cdots \int_V f(y(x))\, dy^1 \wedge \cdots \wedge dy^n = \int \cdots \int_U f(x)\, J\, dx^1 \wedge \cdots \wedge dx^n. \tag{9.12}$$

The sign $\wedge$ means that $dx^i \wedge dx^j = -dx^j \wedge dx^i$, hence under the change of variables $x = x(y)$ we have $dy^i = \frac{\partial y^i}{\partial x^j}\, dx^j$, $dy^i \wedge dy^j = -dy^j \wedge dy^i$ and

finally, as a consequence of these formulas,

$$dy^{i_1} \wedge \cdots \wedge dy^{i_k} = \sum_{j_1 < \cdots < j_k} J^{i_1 \ldots i_k}_{j_1 \ldots j_k} \, dx^{j_1} \wedge \cdots \wedge dx^{j_k},$$

where $J^{(i)}_{(j)}$ is the minor of the Jacobi matrix $\frac{\partial x^i}{\partial y^j}$. For $k = n$ we obtain

$$dy^1 \wedge \cdots \wedge dy^n = J \, dx^1 \wedge \cdots \wedge dx^n,$$

where $J$ is the Jacobian (the determinant of the Jacobi matrix).

Recall that the expression $T \, dx^1 \wedge \cdots \wedge dx^n$ specifies a skew-symmetric tensor of type $(0, n)$, since under coordinate changes it transforms by the rule

$$T \, dx^1 \wedge \cdots \wedge dx^n = T \det\Bigl(\frac{\partial x^i}{\partial y^j}\Bigr) \, dy^1 \wedge \cdots \wedge dy^n = T J^{-1} \, dy^1 \wedge \cdots \wedge dy^n.$$

Another integral considered in classical analysis is the integral of a function over a surface in space. Let $r \colon U \to \mathbb{R}^3$ be a regular imbedding of a surface with coordinates $x^1, x^2$ running over a domain $U \subset \mathbb{R}^2$ and let $f(x^1, x^2)$ be a function on the surface. Then the integral of this function over the surface is defined as the iterated integral

$$\iint_U f(x^1, x^2) \sqrt{g} \, dx^1 \wedge dx^2, \tag{9.13}$$

where $g = \det(g_{ij})$ and $g_{ij} \, dx^i \, dx^j$ is the first fundamental form on the surface.

Note that if an $n$-dimensional manifold is endowed with a Riemannian metric $g_{ij} \, dx^i \, dx^j$, then the metric tensor transforms under coordinate changes $y = y(x)$ by the formula

$$\tilde{g}_{ij} \, dy^i \, dy^j = \tilde{g}_{ij} \frac{\partial y^i}{\partial x^k} \frac{\partial y^j}{\partial x^l} \, dx^k \, dx^l = g_{kl} \, dx^k \, dx^l,$$

so that

$$g = \det(g_{kl}) = \det(\tilde{g}_{ij}) \det\Bigl(\frac{\partial y^i}{\partial x^k}\Bigr) \det\Bigl(\frac{\partial y^j}{\partial x^l}\Bigr) = \tilde{g} J^2.$$

Therefore, under coordinate changes with positive Jacobian, the quantity $\sqrt{g}$ behaves like a tensor of type $(0, n)$.

We can draw the following conclusions from (9.12) and (9.13).

1) Let $U$ be a bounded domain in $n$-dimensional space, and let $T$ be a scalar function of a point in $U$ which behaves like a skew-symmetric tensor of type $(0, n)$ with positive Jacobian. Then the integral

$$\int \cdots \int_U T$$

is well defined.

2) A skew-symmetric tensor $T$ of type $(0, n)$ is written in coordinates as

$$T = T_{1\dots n}\, dx^1 \wedge \cdots \wedge dx^n,$$

and under a change of coordinates $y = y(x)$ (with positive Jacobian) we have

$$\int \cdots \int_V T_{1\dots n}(y(x))\, dy^1 \wedge \cdots \wedge dy^n = \int \cdots \int_U T_{1\dots n}(x) J\, dx^1 \wedge \cdots \wedge dx^n.$$

3) In order to integrate a function $f(x)$ over some domain, we must specify a skew-symmetric tensor $T$ of type $(0, n)$ in this domain (if $T_{1\dots n}$ is everywhere positive, then this tensor is called a volume element, or a *measure*) and the integral of the function $f(x)$ over the domain $U$ is defined as the integral of the tensor $f(x)T$:

$$\int \cdots \int_U f(x) T_{1\dots n}(x)\, dx^1 \wedge \cdots \wedge dx^n.$$

4) If there is a Riemannian metric $g_{ij}$ in the domain $U$, then one can choose for the measure the tensor

$$T = d\sigma = \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n,$$

which in Euclidean coordinates has the form $dx^1 \wedge \cdots \wedge dx^n$.

Thus we obtain the definition of the integral of the first kind, which requires specification of the volume element. In this case we must integrate the function multiplied by this element. If we restrict ourselves to coordinate changes with positive Jacobian, then this integral reduces to an integral of a tensor.

The *integral of the first kind* of a function $f(x)$ over an $n$-dimensional domain $U$ with Riemannian metric $g_{ij}\, dx^i\, dx^j$ is the iterated integral

$$\int \cdots \int_U f(x^1, \dots, x^n) \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n.$$

The integral of the second kind is defined regardless of the choice of a metric, and it is an integral of a skew-symmetric tensor. To each skew-symmetric tensor of type $(0, k)$ corresponds a unique differential $k$-form. The very notion of the differential form is related to the fact that a $k$-form can be integrated over any $k$-dimensional surface.

Let $T$ be a skew-symmetric tensor of type $(0, k)$ in an $n$-dimensional domain $U$, let

$$\sum_{i_1 < \cdots < i_k} T_{i_1 \dots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$$

be the differential form corresponding to this tensor, and let $M^k$ be a regular $k$-dimensional surface in $U$ (i.e., a submanifold) with coordinates $z^1, \dots, z^k$.

Then the integral of the skew-symmetric tensor $T$ of type $(0,k)$ over the $k$-dimensional surface in $U$ (*integral of the second kind*) is defined as

$$\int_{M^k} T = \int \dots \int T'_{1\dots k}\, dz^1 \wedge \dots \wedge dz^k,$$

where the right-hand side is the usual iterated integral of the restriction of the tensor $T$ to the surface (see 8.2.2):

$$\begin{aligned}
&\sum_{i_1<\dots<i_k} T_{i_1\dots i_k}\, dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
&\quad = \sum_{i_1<\dots<i_k} T_{i_1\dots i_k} \frac{\partial x^{i_1}}{\partial z^{j_1}} \cdots \frac{\partial x^{i_k}}{\partial z^{j_k}}\, dz^{j_1} \wedge \dots \wedge dz^{j_k} \\
&\quad = T'_{1\dots k}\, dz^1 \wedge \dots \wedge dz^k.
\end{aligned}$$

For $k = n$ we obtain the definition of the integral of $T$ over a domain $U$.

**Lemma 9.1.** *The value of the integral of the second kind of $T$ over a $k$-dimensional surface in a domain $U$ is not affected by a change of coordinates on the surface (provided that the Jacobian of the transition matrix is positive) and does not depend on the choice of coordinates in $U$.*

**Proof.** If $u^1, \dots, u^k$ are new coordinates on the surface, then the restriction of $T$ to the surface is again a tensor of type $(0,k)$,

$$\tilde{T}'_{1\dots k}\, du^1 \wedge \dots \wedge du^k = \tilde{T}'_{1\dots k} J\, dx^1 \wedge \dots \wedge dx^k = T'_{1\dots k}\, dx^1 \wedge \dots \wedge dx^k,$$

where $J = \det\left(\frac{\partial u^i}{\partial x^j}\right) > 0$, and we have

$$\int_M T = \int_M T'_{1\dots k}\, dx^1 \wedge \dots \wedge dx^k = \int_M \tilde{T}'_{1\dots k}\, du^1 \wedge \dots \wedge du^k.$$

The restriction of $T$ to the surface is not affected by changes of coordinates in the domain $U$. Indeed, if $y = y(x)$, then

$$\begin{aligned}
&\sum_{i_1<\dots<i_k} T_{i_1\dots i_k}\, dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
&\quad = \sum_{i_1<\dots<i_k} T_{i_1\dots i_k} \frac{\partial x^{i_1}}{\partial y^{j_1}} \cdots \frac{\partial x^{i_k}}{\partial y^{j_k}}\, dy^{j_1} \wedge \dots \wedge dy^{j_k} \\
&\quad = \sum_{j_1<\dots<j_k} \tilde{T}_{j_1\dots j_k}\, dy^{j_1} \wedge \dots \wedge dy^{j_k},
\end{aligned}$$

and we have

$$\begin{aligned}
T'_{1\ldots k}\, dz^1 \wedge \cdots \wedge dz^k &= \sum_{i_1<\cdots<i_k} T_{i_1\ldots i_k} \frac{\partial x^{i_1}}{\partial z^{j_1}} \cdots \frac{\partial x^{i_k}}{\partial z^{j_k}}\, dz^{j_1} \wedge \cdots \wedge dz^{j_k} \\
&= \sum_{i_1<\cdots<i_k} T_{i_1\ldots i_k} \Big(\frac{\partial x^{i_1}}{\partial y^{m_1}} \frac{\partial y^{m_1}}{\partial z^{j_1}}\Big) \cdots \Big(\frac{\partial x^{i_k}}{\partial y^{m_k}} \frac{\partial y^{m_k}}{\partial z^{j_k}}\Big)\, dz^{j_1} \wedge \cdots \wedge dz^{j_k} \\
&= \sum_{j_1<\cdots<j_k} \tilde{T}_{m_1\ldots m_k} \frac{\partial y^{m_1}}{\partial z^{j_1}} \cdots \frac{\partial y^{m_k}}{\partial z^{j_k}}\, dz^{j_1} \wedge \cdots \wedge dz^{j_k},
\end{aligned}$$

which proves the lemma. □

EXAMPLE. Let $x(t) = (x^1(t), \ldots, x^n(t))$, $t \in [a, b]$, be a smooth curve, and $T = T_i\, dx^i$ a covector field in the space. The integral (of the first kind) of the field $T$ over the curve $x(t)$ is

$$\int_{x(t)} T_i\, dx^i = \int_a^b T_i(x(t))\, \frac{dx^i(t)}{dt}\, dt.$$

To find the integral over the curve of a scalar function $f(t)$ defined on this curve (e.g., of its curvature), we must introduce the length (volume) element on the curve as $|\dot{x}|\, dt = \sqrt{g_{11}}\, dt = dl$, where $l$ is the natural parameter, and then calculate the integral (of the second kind)

$$\int_a^b f(t)|\dot{x}|\, dt = \int_0^L f(t(l))\, dl,$$

where $L$ is the length of the curve.

REMARK ON THE MOST GENERAL INTEGRAL. The natural question to ask is which integrals do not depend on coordinates, similarly to the integrals of differential forms.

It is not hard to answer this question. Suppose we are given an arbitrary function $F(x, P)$, where $x$ is a point, $P$ is an arbitrary $k$-dimensional tangent plane at this point, and in this plane an oriented $k$-measure is chosen, i.e., an element of the $k$th exterior power of this plane. When this measure is multiplied by a number, $P \to \lambda P$, we require that $F \to \lambda F$ or $F \to |\lambda| F$. Such a quantity may be integrated over any oriented $k$-dimensional surface like a differential form. Briefly, the restriction of this quantity to the surface is a standard object of integration of the same kind as before. The simplest example for any curve is the length element generated by a Riemannian metric $g_{ij}(x)$ (in this case, $P$ is the tangent vector and $F$ is its length). In the same way the Riemannian metric generates the $k$-area of any $k$-dimensional surface. All objects of this kind admit invariant integration.

*The differential forms are singled out by the property that for them, besides integrals, the differentials are defined, which together with integrals possess remarkable topological properties expressed by Stokes-type formulas.*

This is what determines the special role of skew-symmetric tensors. Of particular interest is the extension of the theory of integration to supermanifolds.

**9.2.2. Integral of a form over a manifold.** When defining integrals of the first and second kind in 9.2.1, we used coordinates in the domain over which the integral was taken. At the same time we pointed out that the value of the integral does not depend on the choice of these coordinates (Lemma 9.1). Moreover, the integral is additive:

1) If $U_1$ and $U_2$ are two disjoint domains in which the form $T$ is defined, then

$$\int_{U_1 \cup U_2} T = \int_{U_1} T + \int_{U_2} T.$$

2) If $T_1$ and $T_2$ are skew-symmetric tensors restricted to the surface $M$, then

$$\int_M (T_1 + T_2) = \int_M T_1 + \int_M T_2.$$

We will use these properties to define the integral of a form over a manifold on which there are no global coordinates. The simplest example of such a manifold is a sphere, which may be represented as a union of two pieces with global coordinates.

In the classical calculus, the iterated integral does not depend on the order in which integration is performed:

$$\int dx \int dy\, f(x, y) = \int dy \int dx\, f(x, y).$$

At the same time, interchanging the coordinates $x \leftrightarrow y$ takes a skew-symmetric tensor $f\, dx \wedge dy$ into $-f\, dy \wedge dx$, since the Jacobian of this change is negative. Therefore, in the local definition of the integral of a form as an iterated integral, we actually choose an orientation on the manifold: we set

$$\iint_U f(x, y)\, dx \wedge dy = \int dx \int dy\, f(x, y)$$

when the frames $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ are positively oriented in the domain $U$ (see 5.1.3).

Since the statement of Lemma 9.1 that the integral is not affected by coordinate changes is valid only for orientation-preserving changes, to define integration over a manifold $M^n$ we must require this manifold to be oriented.

Let $M^n$ be an oriented smooth manifold of dimension $n$ on which a differential $n$-form $\omega$ and a partition of unity (see 5.1.5) are defined. Recall that by a partition of unity we mean a covering of the manifold $M^n$ by coordinate domains $U_\alpha$ and a collection of smooth functions $\varphi_\alpha$ with the following properties:

1) Each function $\varphi_\alpha$ is identically equal to zero outside the domain $U_\alpha$ and takes values in the interval $[0,1]$, $0 \le \varphi_\alpha(x) \le 1$.

2) At each point $x \in M^n$ there are only finitely many nonzero functions $\varphi_\alpha$ and

$$\sum_\alpha \varphi_\alpha(x) \equiv 1.$$

For simplicity we will consider the case where the partition of unity consists of finitely many domains $U_\alpha$. Such a partition always exists if the manifold is compact.

The *integral* (of the second kind) of a form $T = T_{1\dots n}\, dx^1 \wedge \dots \wedge dx^n$ over the manifold $M^n$ is the quantity

$$\int_{M^n} T = \sum_\alpha \int_{U_\alpha} \varphi_\alpha T,$$

where the summands $\int_{U_\alpha} \varphi_\alpha T$ are integrals (of the second kind), defined as iterated integrals, of the forms $\varphi_\alpha T$ over the domains $U_\alpha$. This means that

$$\int_{U_\alpha} \varphi_\alpha T = \int_{U_\alpha} \varphi_\alpha T_{1\dots n}\, dx_\alpha^1 \cdots dx_\alpha^n,$$

where the bases consisting of tangent vectors $\left(\frac{\partial}{\partial x_\alpha^1}, \dots, \frac{\partial}{\partial x_\alpha^n}\right)$ are positively oriented in the corresponding domains $U_\alpha$.

The following simple argument shows that this definition does not depend on the choice of the partition of unity. Let $x$ be a point lying together with its neighborhood $U$ in the intersection of domains $U_1, \dots, U_k$. Since the value of the integral over a coordinate domain does not depend on the choice of coordinates, we obviously have

$$\sum_{i=1}^k \int_U \varphi_i T = \int_U \left(\sum_{i=1}^k \varphi_i\right) T = \int_U T.$$

Therefore

$$\int_{M^n} \omega = \int_{M^n \setminus U} T + \int_U T.$$

The value of the second integral does not depend on the choice of the partition of unity, thus the value of the second kind integral taken over a neighborhood of any point does not depend on the choice of the partition of unity. Hence it is clear that the integral over the entire manifold is also well defined.

The *integral* (of the first kind) of a function $f(x)$ on an oriented manifold $M^n$ with Riemannian metric $g_{ij}$ is defined to be

$$\int_{M^n} f\, d\sigma = \int_{M^n} f\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n,$$

where the integral of $f\, d\sigma$ is defined with the aid of a partition of unity as the integral of a skew-symmetric tensor.

We point out one more possibility to define the integral over a manifold without using a partition of unity. On some manifolds one can introduce global coordinates with singularities lying in a set of smaller dimension, such as, e.g., spherical coordinates on spheres (for a two-dimensional sphere they were defined in 1.1.2) or polar coordinates on the plane. The integrals of forms of maximal degree over the sets of singular points vanish, and such coordinates are often used in the theory of integration.

**9.2.3. Integrals of differential forms in $\mathbb{R}^3$.** Consider integrals of differential forms over submanifolds in Euclidean space $\mathbb{R}^3$ with Euclidean coordinates $x^1, x^2, x^3$.

1) The integral of a tensor of rank zero, i.e., of a scalar $f(x)$, over a 0-dimensional submanifold (a point $P$) is simply the value of the function $f(x)$ at the point $P$, i.e., the integral equals $f(P)$.

2) The integral of a tensor of rank 1, i.e., of a covector field $T_i\, dx^i$ over a curve $\Gamma$ specified as $x(t) = (x^1(t), x^2(t), x^3(t))$, $a \le t \le b$, is

$$\int_\Gamma T_i\, dx^i = \int_a^b T_i(x(t))\dot{x}^i\, dt.$$

If the curve is closed, $x(a) = x(b)$, this integral,

$$\oint_\Gamma T_i\, dx^i,$$

is called the *circulation of the field* $T$ along the curve $\Gamma$. Since vectors and covectors are not distinguished in Euclidean coordinates, this integral is usually referred to as the circulation of the vector field $T = (T_1, T_2, T_3)$.

3) In general, the integral of a tensor field $(T_{ij}) = \sum_{i<j} T_{ij}\, dx^i \wedge dx^j$ over a two-dimensional surface $\Gamma$ in $\mathbb{R}^n$ specified by equations $x^i = x^i(z^1, z^2)$, $i = 1, \dots, n$, is

$$\begin{aligned}\int_\Gamma \sum_{i<j} T_{ij}\, dx^i \wedge dx^j &= \int_\Gamma \sum_{i<j} T_{ij}(x(z))\Big(\frac{\partial x^i}{\partial z^k}\, dz^k\Big) \wedge \Big(\frac{\partial x^j}{\partial z^l}\, dz^l\Big) \\ &= \int_\Gamma \Big[\sum_{i<j} T_{ij}\Big(\frac{\partial x^i}{\partial z^1}\frac{\partial x^j}{\partial z^2} - \frac{\partial x^j}{\partial z^1}\frac{\partial x^i}{\partial z^2}\Big)\Big]\, dz^1 \wedge dz^2.\end{aligned} \tag{9.14}$$

At each point of the surface, the tangent vectors

$$\xi = \left(\frac{\partial x^i}{\partial z^1}\right), \quad \eta = \left(\frac{\partial x^i}{\partial z^2}\right)$$

constitute a basis in the tangent space at this point. In three-dimensional Euclidean space the vector product of these vectors

$$[\xi, \eta] = \det \begin{pmatrix} e_1 & e_2 & e_3 \\ \xi^1 & \xi^2 & \xi^3 \\ \eta^1 & \eta^2 & \eta^3 \end{pmatrix}$$

is orthogonal to the surface (here $e_1, e_2, e_3$ are an orthonormal basis in $\mathbb{R}^3$). The length of the vector $[\xi, \eta]$ is equal to $\sqrt{g}$, where $g = \det(g_{ij})$ is the determinant of the first fundamental form of the surface (see 3.2.1),

$$g_{ij}\, dz^i\, dz^j = \sum_{i=1}^{3} (dx^i)^2, \quad dx^i = \frac{\partial x^i}{\partial z^j}\, dz^j.$$

To a skew-symmetric tensor $T$ of rank 2 there corresponds the covector field $*T$, which will be denoted by $\tilde{T}$ and which in Euclidean coordinates may be regarded as the vector field

$$\tilde{T} = (\tilde{T}^1, \tilde{T}^2, \tilde{T}^3) = (T_{23}, -T_{13}, T_{12}).$$

Substituting the expressions for $T$ and $[\xi, \eta]$ into (9.14) we obtain

$$(9.15) \quad \int_\Gamma \sum_{i<j} T_{ij}\, dx^i \wedge dx^j = \int_\Gamma \langle \tilde{T}, [\xi, \eta] \rangle\, dz^1 \wedge dz^2 = \int_\Gamma \langle \tilde{T}, \mathbf{n} \rangle \sqrt{g}\, dz^1 \wedge dz^2,$$

where $\mathbf{n}$ is the unit normal vector to the surface, $\mathbf{n} = \frac{[\xi,\eta]}{|[\xi,\eta]|} = \frac{[\xi,\eta]}{\sqrt{g}}$.

Thus we have proved the following theorem.

**Theorem 9.6.** *In the three-dimensional space $\mathbb{R}^3$ the integral of a 2-form $T$ over a surface coincides with the integral of the first kind of the function $\langle \tilde{T}, \mathbf{n} \rangle$, where $\mathbf{n}$ is the normal to the surface and $\tilde{T} = *T$.*

In spaces of dimension greater than 3, the integrals of 2-forms cannot be reduced to operations only on vector fields, since the operator $*$ does not take 2-forms into covector fields.

If the surface $\Gamma$ is closed (is the boundary of some domain), then the integral

$$\int_\Gamma \sum_{i<j} T_{ij}\, dx^i \wedge dx^j$$

is called the *total flux of the field* $\tilde{T}$ over the surface $\Gamma$.

**9.2.4. Stokes theorem.** In the classical calculus there are a number of results that relate integrals over a domain and its boundary (the Green, Gauss–Ostrogradskii, and Stokes formulas). All of them are special cases of the Stokes theorem.

The simplest connection between integrals over a domain and its boundary is established by the Newton–Leibniz formula: if $f(x)$ is a smooth function on a curve $M$ specified by the equation $x = x(t)$, $a \leq t \leq b$, then

$$\int_M df = \int_M \frac{\partial f}{\partial x^i} \dot{x}^i \, dt = f(x(b)) - f(x(a)).$$

Here the pair of endpoints $P = x(a)$ and $Q = x(b)$ may be regarded as the boundary of the one-dimensional manifold $M$ with orientation determined by the direction of the curve (parameter $t$ varies from $a$ to $b$). We will formally write the boundary of $M$ as

$$\partial M = Q - P.$$

Then the Newton–Leibniz formula becomes

$$\int_M df = \int_{\partial M} f,$$

where $f$ is a 0-form (a scalar function).

Consider now the general case.

Let $M^n$ be an oriented smooth manifold with smooth boundary $\partial M^n$ (possibly, empty). Define the orientation on the boundary by the following rule: a basis $e_1, \ldots, e_{n-1}$ of tangent vectors to the boundary at a point $x \in \partial M^n$ is positively oriented if the basis $v, e_1, \ldots, e_{n-1}$ is positively oriented, where $w = -v$ is an inward normal to $\partial M^n$ at the point $x$ (the vector $v$ is called an outward normal).

As an example of such a manifold we may consider a domain $U$ in $\mathbb{R}^n$ with smooth boundary $\Gamma$.

**Theorem 9.7** (Stokes). *Let $T$ be an $(n-1)$-form on an oriented compact smooth manifold of dimension $n$ with smooth or piecewise-smooth boundary $\partial M^n$. Then*

$$\int_{\partial M^n} T = \int_{M^n} dT. \tag{9.16}$$

*In particular, if the manifold $M^n$ is closed (has no boundary), then the integral of any form $dT$ over this manifold is equal to zero.*

The boundary is said to be piecewise-smooth if it consists of finitely many smooth pieces which mutual boundaries of smaller dimension. For example, the Cartesian cube

$$I^n = \{(x^1, \ldots, x^n) \colon 0 \leq x^i \leq 1,\ i = 1, \ldots, n\}$$

in $\mathbb{R}^n$ is piecewise smooth. By definition, the integral over a piecewise-smooth boundary is the sum of the integrals over its smooth pieces.

Consider some particular cases of the general Stokes formula (9.16) familiar to the reader from a course of calculus.

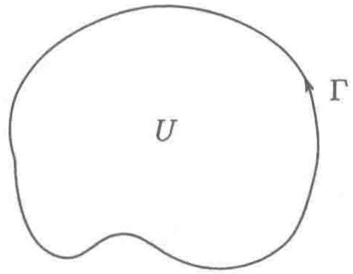


**Figure 9.1.** Orientation of the boundary of a domain.

EXAMPLES. 1) GREEN'S FORMULA. Consider a closed contour $\Gamma$ that bounds a domain $U$ on the plane with coordinates $x, y$ (Figure 9.1). Then, for a smooth covector field $T = T_1\,dx + T_2\,dy$, the following Green formula (a particular case of (9.16) with $M^2 = U$) holds:

$$\int_\Gamma (T_1\,dx + T_2\,dy) = \iint_U \Big(\frac{\partial T_2}{\partial x} - \frac{\partial T_1}{\partial y}\Big)\,dx \wedge dy. \tag{9.17}$$

Here integration over $\Gamma$ goes counterclockwise.

2) CAUCHY'S FORMULA. We apply the Green formula to a smooth complex-valued function $f(z) = f(x,y) = u(x,y)+iv(x,y)$, where $z = x+iy$. We have

$$\begin{aligned}\oint_\Gamma f(z)\,dz &= \oint_\Gamma (u+iv)(dx+idy) = \oint_\Gamma (u\,dx - v\,dy) + i\oint_\Gamma (v\,dx + u\,dy)\\ &= -\iint_U \Big(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\Big)\,dx\wedge dy + i\iint_U \Big(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\Big)\,dx\wedge dy.\end{aligned}$$

Therefore, if $f(z)$ is complex-analytic in $U$, or, equivalently, if it satisfies the Cauchy–Riemann conditions (see 4.1.4),

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \quad \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x},$$

then $\oint_\Gamma f(z)\,dz = 0$ and the form $f(z)\,dz$ is closed in $U$.

If $f(z)$ is complex-analytic in $U$, then the integral of $f(z)/(z-a)$ over $\Gamma$, where $a$ is an interior point of $U$, is equal to

$$\frac{1}{2\pi i}\oint_\Gamma \frac{f(z)}{z-a}\,dz = f(a).$$

This is the *residue formula* or *Cauchy's formula*. To prove it, we delete from $U$ a ball $B$ of radius $\varepsilon$ with center at $a$. Let $\Gamma'$ be the boundary of this ball. Then (9.16) implies that

$$\oint_\Gamma \frac{f(z)}{z-a}\,dz - \oint_{\Gamma'} \frac{f(z)}{z-a}\,dz = \int_{U\setminus B} d\Big(\frac{f(z)}{z-a}\Big) = 0.$$

The integral over $\Gamma'$ is taken with the opposite sign to make the orientation of this boundary contour compatible with that of $U$. Let us write down the Taylor series expansion of $f(z)$ about $z = a$:

$$f(z) = f(a) + \sum_{k\geq 1} c_k (z-a)^k.$$

For the circle $\Gamma' = \{a + \varepsilon e^{it}\}$ we obtain by a direct calculation that

$$\oint_{\Gamma'} (z-a)^n\,dz = \begin{cases} 0 & \text{if } n \neq -1, \\ 2\pi i & \text{if } n = -1. \end{cases}$$

Substituting now the Taylor series for $f(z)$ into the integral $\oint_{\Gamma'} \frac{f(z)}{z-a}\,dz$, we obtain

$$\oint_\Gamma \frac{f(z)}{z-a}\,dz = \oint_{\Gamma'} \frac{f(z)}{z-a}\,dz = 2\pi i f(a),$$

which proves Cauchy's formula.

For a Laurent series $f(z) = \sum_{k\geq k_0 > -\infty} c_k (z-a)^k$ in the domain $U$ (which for $k_0 \geq 0$ turns into the Taylor series for a complex-analytic function), Cauchy's formula enables us to compute all the coefficients of the expansion:

$$\oint_\Gamma (z-a)^{-k} f(z)\,dz = 2\pi i c_{k-1}.$$

3) The Gauss–Ostrogradskii formula. Let $U$ be a domain in $\mathbb{R}^3$ with Euclidean coordinates $x^1, x^2, x^3$, and let $\Gamma$ be its piecewise-smooth boundary. For a tensor field $T = \sum_{i<j} T_{ij}\,dx^i \wedge dx^j$, formula (9.16) implies that

$$\text{(9.18)}\quad \iint_\Gamma \sum_{i<j} T_{ij}\,dx^i \wedge dx^j = \iiint_U \Big(\frac{\partial T_{12}}{\partial x^3} + \frac{\partial T_{23}}{\partial x^1} - \frac{\partial T_{13}}{\partial x^2}\Big)\,dx^1 \wedge dx^2 \wedge dx^3.$$

We rewrite this formula in terms of the vector field $T = (T_{23}, -T_{13}, T_{12})$. By (9.15) we obtain

$$\iint_\Gamma \sum_{i<j} T_{ij}\,dx^i \wedge dx^j = \iint_\Gamma \langle T, \mathbf{n}\rangle \sqrt{g}\,dz^1 \wedge dz^2 = \iint_\Gamma \langle T, \mathbf{n}\rangle\,d\sigma,$$

where $z^1, z^2$ are coordinates on the surface and $d\sigma = \sqrt{g}\,dz^1 \wedge dz^2$ is the area element. Note that

$$\frac{\partial T_{12}}{\partial x^3} - \frac{\partial T_{13}}{\partial x^2} + \frac{\partial T_{23}}{\partial x^1} = \frac{\partial T^i}{\partial x^i} = \operatorname{div} T.$$

Substituting these expressions into (9.18), we obtain the *Gauss–Ostrogradskii* (or *divergence*) formula:

$$\iint_{\Gamma} \langle T, \mathbf{n} \rangle \, d\sigma = \iiint_U (\operatorname{div} T) \, dx^1 \wedge dx^2 \wedge dx^3. \tag{9.19}$$

4) THE STOKES FORMULA. Let $U$ be a domain on a surface $x^i = x^i(z^1, z^2)$, $i = 1, 2, 3$, in $\mathbb{R}^3$, and let $\Gamma$ be the boundary of this domain. For a covector field $T_i \, dx^i$ we have

$$\oint_{\Gamma} T_i \, dx^i = \iint_U \Big( \Big( \frac{\partial T_2}{\partial x^1} - \frac{\partial T_1}{\partial x^2} \Big) dx^1 \wedge dx^2 + \Big( \frac{\partial T_3}{\partial x^1} - \frac{\partial T_1}{\partial x^3} \Big) dx^1 \wedge dx^3 + \Big( \frac{\partial T_3}{\partial x^2} - \frac{\partial T_2}{\partial x^3} \Big) dx^2 \wedge dx^3 \Big).$$

Now (9.15) implies the *Stokes formula*

$$\oint_{\Gamma} T_\alpha \, dx^\alpha = \iint_U \langle \operatorname{curl} T, \mathbf{n} \rangle \sqrt{g} \, dz^1 \wedge dz^2. \tag{9.20}$$

Note that in expressions such as $\int_M \langle T, \mathbf{n} \rangle \, dz^1 \wedge dz^2$, the choice of the sign (orientation of the boundary) is determined by the direction of the normal $\mathbf{n}$.

Now we indicate some important physical applications of the general Stokes formula (9.16).

Recall that the Maxwell equations of electromagnetic field break up into two pairs of equations for the vectors $\mathbf{E} = (E_1, E_2, E_3)$ and $\mathbf{H} = (H_1, H_2, H_3)$ of electric and magnetic fields in three-dimensional Euclidian space (see 9.1.3).

The first pair of equations is

$$\operatorname{div} \mathbf{H} = \frac{\partial H^i}{\partial x^i} = 0, \tag{9.21}$$

$$\operatorname{curl} \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} = 0, \tag{9.22}$$

where $c$ is the velocity of light in vacuum.

From (9.21) and the Gauss–Ostrogradskii formula (9.19) we obtain

$$\iiint_U \operatorname{div} \mathbf{H} \, dx^1 \wedge dx^2 \wedge dx^3 = \iint_{\partial U} \langle \mathbf{H}, \mathbf{n} \rangle \, d\sigma = 0,$$

which implies that:

*The flux of a magnetic field through a closed surface is equal to zero.*

Equation (9.22) and the Stokes formula (9.20) imply that

$$-\frac{1}{c} \iint_U \Big\langle \frac{\partial \mathbf{H}}{\partial t}, \mathbf{n} \Big\rangle \, d\sigma = \iint_U \langle \operatorname{curl} \mathbf{E}, \mathbf{n} \rangle \, d\sigma = \oint_{\partial U} E_\alpha \, dx^\alpha,$$

which says that

*The derivative with respect to time of the flux of a magnetic field through a closed surface is equal, up to the factor* $-\frac{1}{c}$, *to the circulation of the electric field over the boundary of the surface.*

The second pair of the Maxwell equations in the three-dimensional case is written as

$$\operatorname{div} \mathbf{E} = 4\pi\rho, \tag{9.23}$$

$$\operatorname{curl} \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{j}, \tag{9.24}$$

where $\rho$ is the density of charge and $\mathbf{j}$ is the vector of current.

Now (9.23) and the Gauss–Ostrogradskii formula (9.19) imply that

$$\iiint_U 4\pi\rho\, dx^1 \wedge dx^2 \wedge dx^3 = \iint_{\partial U} \langle \mathbf{E}, \mathbf{n} \rangle\, d\sigma.$$

This means that

*The flux of an electric field through the boundary of a domain in space is equal, up to* $4\pi$, *to the total charge in this domain.*

From the Stokes formula (9.20) and (9.24) we obtain

$$\frac{1}{c}\iint_U \Big\langle \frac{\partial \mathbf{E}}{\partial t}, \mathbf{n} \Big\rangle\, d\sigma + \frac{4\pi}{c}\iint_U \langle \mathbf{j}, \mathbf{n} \rangle\, d\sigma = \iint_U \langle \operatorname{curl} \mathbf{H}, \mathbf{n} \rangle\, d\sigma = \oint_{\partial U} H_\alpha\, dx^\alpha.$$

The physical meaning of this equality is as follows:

*The circulation of the magnetic field along the boundary of a surface is equal to the sum of the net current through this surface and the derivative with respect to time of the flux of the electric field through this surface.*

**9.2.5. The proof of the Stokes theorem for a cube.** We will prove the Stokes theorem only for a cube. The idea of the proof (reduction to the Newton–Leibniz formula) will be clear. In fact, the general case is obtained from the proof for a cube by additivity.

A *singular $k$-dimensional cube* $\sigma$ in $\mathbb{R}^n$ is a smooth mapping $\sigma\colon I^k \to \mathbb{R}^n$, where $I^k$ is the $k$-dimensional Cartesian unit cube,

$$I^k = \{(x^1, \dots, x^k)\colon 0 \le x^\alpha \le 1\}.$$

The equations $x^j = 0$ and $x^i = 1$ determine two $(k-1)$-faces $l_j^-$ and $l_j^+$, respectively.

According to the convention about orientation of the boundary adopted in 9.2.4, we endow the faces $l_j^-$ and $l_j^+$ with the signs $(-1)^j$ and $(-1)^{j+1}$, respectively (i.e., the orientation is determined by the outward normal).

Let $\varphi$ be a $(k-1)$-form in $\mathbb{R}^n$, and $d\varphi$ its exterior differential.

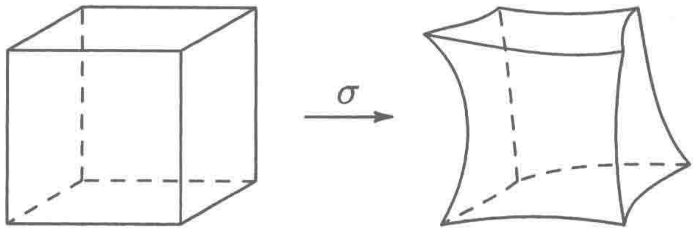

**Figure 9.2.** Singular cube.

**Theorem 9.8.** *For a singular $k$-dimensional cube $I^k$ in $\mathbb{R}^n$,*

$$\int_{\sigma(\partial I^k)} \varphi = \int_{\sigma(I^k)} d\varphi.$$

**Proof.** Let $\omega = \sigma^*(\varphi)$. Since the operations $\sigma$ and $d$ commute, we have $d\omega = \sigma^*(d\varphi)$. By definition,

$$\int_{\sigma(I^k)} d\varphi = \int_{I^k} \sigma^*(d\varphi), \quad \int_{\sigma(\partial I^k)} \varphi = \int_{\partial I^k} \sigma^*(\varphi);$$

hence we must prove that

$$\int_{\partial I^k} \omega = \int_{I^k} d\omega.$$

The form $\omega$ is written as a sum

$$\omega = \sum_i T_i(x^1, \dots, x^k)\, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k,$$

where $T_i(x^1, \dots, x^k)$ are smooth functions. By the definition of the exterior differential, we have

$$d\omega = \sum_i \frac{\partial T_i}{\partial x^i}\, dx^i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k = \sum_i (-1)^{i-1} \frac{\partial T_i}{\partial x^i}\, d^k x,$$

where $d^k x = dx^1 \wedge \cdots \wedge dx^k$.

By a direct calculation we obtain

$$\begin{aligned}
\int_{I^k} d\omega &= \int_{I^k} \Big( \sum_i (-1)^{i-1} \frac{\partial T_i}{\partial x^i} \Big)\, d^k x \\
&= \sum_i \int_{(x^1)} \cdots \int_{(\widehat{x^i})} \cdots \int_{(x^k)} \Big[ \int_{(x^i)} \frac{\partial T_i}{\partial x^i}\, dx^i \Big]\, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k \\
&= \sum_i \int_{(x^1)} \cdots \int_{(\widehat{x^i})} \cdots \int_{(x^k)} (T_i(x^1, \dots, x^k)|_{x^i=1} \\
&\qquad\qquad - T_i(x^1, \dots, x^k)|_{x^i=0})\, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^k \\
&= \int_{\partial I^k} \omega.
\end{aligned}$$

Hence the theorem. □

**9.2.6. Integration over a superspace.** Modern physics considers transformations that mix bosons and fermions (supersymmetries). To deal with them, we need differential and integral calculus on the space $\mathbb{R}^{m|n}$.

In $\mathbb{R}^{m|n}$ there are natural real-valued coordinates $x^1, \dots, x^m$ and anticommuting coordinates $\xi^1, \dots, \xi^n$, which generate the exterior algebra $\Lambda\mathbb{R}^n$. Multiplication in $\mathbb{R}^{m|n}$ satisfies the following commutation relations:

$$x^i x^j = x^j x^i, \quad x^i \xi_j = \xi_j x^i, \quad \xi_i^2 = 0, \quad \xi_i \xi_j = -\xi_j \xi_i \qquad \text{for} \quad i \neq j.$$

Smooth functions on $\mathbb{R}^{m|n}$ have the form of polynomials in anticommuting variables,

$$f(x, \xi) = f_0(x) + \sum_{k=1}^{n} \sum_{i_1 < \cdots < i_k} f_{i_1 \dots i_k} \xi_{i_1} \cdots \xi_{i_k},$$

where the coefficients $f_0$ and $f_{i_1 \dots i_k}$ are smooth functions of $x^1, \dots, x^m$.

Functions and variables are said to be even if they commute with any other functions (in which case $f_{i_1 \dots i_k} = 0$ for odd $k$). Otherwise they are called odd.

The derivatives of $f$ are defined by the rules

$$\begin{aligned}
\frac{\partial f}{\partial x^j} &= \frac{\partial f_0}{\partial x^j} + \sum_{k=1}^{n} \sum_{i_1 < \cdots < i_k} \frac{\partial f_{i_1 \dots i_k}}{\partial x^j} \xi_{i_1} \cdots \xi_{i_k}, \\
\frac{\partial f}{\partial \xi_j} &= h(x, \xi_1, \dots, \widehat{\xi_j}, \dots, \xi_n),
\end{aligned}$$

where

$$f(x, \xi) = g(x, \xi_1, \dots, \widehat{\xi_j}, \dots, \xi_n) + \xi_j \wedge h(x, \xi)$$

(the "hat" means that the expansion of $g$ into a polynomial in anticommuting variables does not contain $\xi_j$).

Consider nonlinear changes of coordinates

$$x'^i = x'^i(x,\xi), \quad \xi'_j = \xi'_j(x,\xi).$$

Each function $x'^i$ itself is represented as a polynomial of anticommuting variables,

$$x'^i = x_0'^i(x^1,\dots,x^m) + \cdots,$$

where the last dots denote the nilpotent part, i.e., the terms which vanish on raising them to some power. In order to define, for any smooth function $f(x^1,\dots,x^m)$, the result of the substitution

$$f(x'^1,\dots,x'^m),$$

we write down the formal Taylor series for $f$ about the point $(x'^1,\dots,x'^m)$ and consider the nilpotent part $(x'-x'_0)$ as the increment $\Delta x$. Since the increment is nilpotent, the Taylor series will be a polynomial in anticommuting variables. For example,

$$e^{x+\xi_1\wedge\xi_2} = e^x + \xi_1\wedge\xi_2,$$
$$\sin(x+\xi_1\wedge\xi_2) = \sin x + \cos x\ \xi_1\wedge\xi_2.$$

The above rules are enough for developing differential calculus on $\mathbb{R}^{m|n}$. Now we proceed to integration.

The definition of integral over anticommuting variables presented in 7.5.2 is extended to integration over the entire space $\mathbb{R}^{m|n}$ by the formula

$$\int_{\mathbb{R}^{m|n}} f(x,\xi)\,dx^1\wedge\cdots\wedge dx^m\wedge d\xi_1\wedge\cdots\wedge d\xi_n = \int_{\mathbb{R}^m} f_{1\dots n}(x)\,dx^1\wedge\cdots\wedge dx^m.$$

Now, let $x' = x'(x,\xi)$, $\xi' = \xi'(x,\xi)$ be a change of variables. Its Jacobi matrix is

$$J = \begin{pmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial \xi} \\ \frac{\partial \xi'}{\partial x} & \frac{\partial \xi'}{\partial \xi} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Since the variables $x'$ are even and $\xi'$ odd, the differentiation rules imply that the matrices $A$ and $D$ consist of even elements, while $B$ and $C$ consist of odd elements.

In Chapter 7 we proposed a problem to prove that such a matrix is invertible if and only if the matrices $A$ and $D$ are invertible. Therefore, $A$ has an inverse $A^{-1}$ and $\det A \det A^{-1} = 1$. Hence we see that the determinant of $A$ has a nonzero numerical part, i.e.,

$$\det A = a_0 + O(\xi^2), \quad \det A^{-1} = a_0^{-1} + O(\xi^2).$$

By definition, we let

$$\operatorname{sgn}\det A = \operatorname{sgn} a_0.$$

A formal calculation of the determinant of $A$ encounters difficulties, since matrix entries have odd components (and so we must take into account the order of variables when multiplying them). A natural analog of the determinant for matrices containing anticommuting variables is defined only in the case where the matrix $D$ is invertible:

$$\operatorname{Ber} J = \frac{\det(A - BD^{-1}C)}{\det D}.$$

It is called the *Berezinian*. Since all the components of $A$ and $BD^{-1}C$ are even (commute with all variables), a formal calculation of this determinant presents no difficulties.

Now integration with respect to any coordinates is reduced to integration with respect to $x^1, \dots, x^m, \xi_1, \dots, \xi_n$ by the formula

$$\int_{\mathbb{R}^{m|n}} f(x', \xi')\, d^m x' \wedge d^n \xi'$$
$$= \operatorname{sgn}\det\left(\frac{\partial x'}{\partial x}\right) \int_{\mathbb{R}^{m|n}} f(x'(x,\xi), \xi'(x,\xi)) \operatorname{Ber} J\, d^m x \wedge d^n \xi.$$

## 9.3. Cohomology

**9.3.1. De Rham cohomology.** Let $M^n$ be a smooth manifold of dimension $n$. Denote by $C^p(M^n)$ the linear space of all smooth $p$-forms on $M^n$. The exterior differentiation determines homomorphisms

$$d\colon C^p(M^n) \to C^{p+1}(M^n),$$

which form a sequence

$$C^0(M^n) \xrightarrow{d} C^1(M^n) \xrightarrow{d} \cdots \xrightarrow{d} C^p(M^n) \xrightarrow{d} C^{p+1}(M^n) \xrightarrow{d} \cdots \xrightarrow{d} C^n(M^n),$$

with the following important property:

$$d^2 = 0. \tag{9.25}$$

A linear mapping $d\colon C \to C$ of an arbitrary linear space $C$ into itself (in our case $C = \bigoplus C^p(M)$) is called a *differential* if it satisfies condition (9.25). The pair $(C, d)$ is called a *complex*.

In any space $C^p(M^n)$ one can single out the following two subspaces:

1) the subspace $Z^p(M^n)$ of all closed $p$-forms (which are called *cocycles*);

2) the subspace $B^p(M^n)$ of all exact $p$-forms, i.e., the exterior derivatives of $(p-1)$-forms (they are called *coboundaries*).

By definition we have

$$Z^p(M^n) = \operatorname{Ker} d \cap C^p(M^n), \quad B^p(M^n) = \operatorname{Im} d \cap C^p(M^n).$$

Formula (9.25) implies that

$$\operatorname{Im} d \subset \operatorname{Ker} d,$$

i.e., all coboundaries are cocycles. The quotient space

$$H^*(M^n; \mathbb{R}) = \operatorname{Ker} d / \operatorname{Im} d$$

is called the *de Rham cohomology* space.

This is a particular case of the following general construction: if we are given a linear space $C$ on which a differential $d\colon C \to C$ acts, then the quotient space $H = \operatorname{Ker} d / \operatorname{Im} d$ is called the *cohomology* space of the complex $(C, d)$.

De Rham cohomology is graded, i.e., it breaks up into the direct sum of $p$-dimensional cohomology spaces $H^p(M^n; \mathbb{R})$:

$$H^*(M^n; \mathbb{R}) = \bigoplus_{p \geq 0} H^p(M^n; \mathbb{R}), \quad H^p(M^n; \mathbb{R}) = Z^p(M^n)/B^p(M^n).$$

The following result is an immediate consequence of definitions.

**Corollary 9.1.** *If $M^n$ is an $n$-dimensional manifold, then $H^p(M^n; \mathbb{R}) = 0$ for $p > n$.*

**Lemma 9.2.** *The 0-dimensional cohomology space $H^0(M^n; \mathbb{R})$ is generated by connected components of the manifold $M^n$.*

**Proof.** The 0-dimensional smooth forms are smooth functions on $M^n$. By definition $B^0(M^n) = 0$. A 0-dimensional form $f$ is closed if its gradient vanishes everywhere, i.e., the function $f$ is constant on each connected component of $M^n$. Therefore, the 0-dimensional cohomology class $[f]$ is uniquely determined by the values of $f$ on connected components of the manifold. Hence the lemma. □

**Corollary 9.2.** *Let $X = \mathrm{pt}$ be a point. Then $H^0(X; \mathbb{R}) = \mathbb{R}$ and $H^p(X; \mathbb{R}) = 0$ for $p > 0$.*

The elements of the cohomology space are the equivalence classes $[\omega]$ of closed modulo coboundaries:

$$[\omega_1] = [\omega_2] \quad \text{if and only if} \quad \omega_1 - \omega_2 = d\varphi.$$

If closed forms $\omega$ and $\omega'$ belong to the same cohomology class, then they are said to be *cohomological* to each other.

The cohomology space is, obviously, an Abelian group with addition

$$[\omega_1] + [\omega_2] = [\omega_1 + \omega_2].$$

This group contains $p$-dimensional cohomology subgroups $H^p(M^n, \mathbb{R})$.

The following lemma shows that this group is endowed also with a multiplication operation.

**Lemma 9.3.** *The operation of multiplication of smooth forms generates multiplication in cohomology,*

$$[\omega_1] \smile [\omega_2] = [\omega_1 \wedge \omega_2],$$

*which is associative and has the following properties:*

$$H^p \smile H^q \subset H^{p+q}, \quad a \smile b = (-1)^{pq} b \smile a \quad \text{for } a \in H^p,\ b \in H^q.$$

**Proof.** We will show that the equivalence class $[\omega_1] \smile [\omega_2]$ does not depend on the choice of the smooth forms $\omega_1$ and $\omega_2$. Then the algebraic properties follow from the properties of multiplication of smooth forms. Let $\omega_1 \in C^p(M^n)$; then

$$\begin{aligned}(\omega_1 + d\varphi_1) \wedge (\omega_2 + d\varphi_2) &= \omega_1 \wedge \omega_2 + \omega_1 \wedge d\varphi_2 + d\varphi_1 \wedge \omega_2 + d\varphi_1 \wedge d\varphi_2 \\ &= \omega_1 \wedge \omega_2 + d((-1)^p \omega_1 \wedge \varphi_2 + \varphi_1 \wedge \omega_2 + \varphi_1 \wedge d\varphi_2) = \omega_1 \wedge \omega_2 + d\psi.\end{aligned}$$

Therefore

$$[(\omega_1 + d\varphi_1) \wedge (\omega_2 + d\varphi_2)] = [\omega_1 \wedge \omega_2].$$

Hence the lemma. □

**Corollary 9.3.** *The cohomology class $H^*(M^n; \mathbb{R})$ with the above addition and multiplication is a ring (real cohomology ring).*

**9.3.2. Homotopy invariance of cohomology.** Let $f\colon M^n \to N^k$ be a smooth mapping. Then it generates a mapping of smooth forms

$$f^*\colon C^p(N^k) \to C^p(M^n), \quad p \geq 0.$$

Since $df^* = f^*d$, the mapping $f^*$ carries closed forms into closed forms and exact forms into exact forms. Therefore, it generates a linear mapping of cohomology rings by the formula

$$f^*\colon H^*(N^k; \mathbb{R}) \to H^*(M^n; \mathbb{R}), \quad f^*[\omega] = [f^*\omega].$$

Furthermore, since $f^*(\omega_1 \wedge \omega_2) = f^*(\omega_1) \wedge f^*(\omega_2)$, this mapping is a homomorphism of cohomology rings:

$$f^*(a \smile b) = f^*a \smile f^*b, \quad a, b \in H^*(N^k; \mathbb{R}).$$

In this case $f^*$ is referred to as the *homomorphism induced by the mapping* $f$.

Recall that two smooth mappings $f_0\colon M^n \to N^k$ and $f_1\colon M^n \to N^k$ are called (smoothly) homotopic if there is a smooth mapping $F\colon M^n \times [0,1] \to N^k$ such that it coincides with the mappings $f_0$ and $f_1$ on the boundaries of the cylinder $M^n \times [0,1]$:

$$F(x, k) = f_k(x), \quad k = 0, 1.$$

This mapping $F$ is called a *homotopy* between the mappings $f_0$ and $f_1$.

**Theorem 9.9.** *If smooth mappings $f_0\colon M^n \to N^k$ and $f_1\colon M^n \to N^k$ are homotopic, then the homomorphisms of cohomology rings induced by them,*

$$f_0^*\colon H^*(N^k;\mathbb{R}) \to H^*(M^n;\mathbb{R}) \quad \textit{and} \quad f_1^*\colon H^*(N^k;\mathbb{R}) \to H^*(M^n;\mathbb{R}),$$

*coincide:*

$$f_0^* = f_1^*.$$

First we will prove a technical lemma.

Any $p$-form $\Omega$ on the cylinder $M^n \times [0,1]$ breaks up into the sum

$$\Omega = \omega_{p-1} \wedge dt + \omega_p,$$

where

$$\omega_{p-1} = \sum_{i_1<\cdots<i_{p-1}} a_{i_1\dots i_{p-1}}(x,t)\, dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}},$$
$$\omega_p = \sum_{j_1<\cdots<j_p} b_{j_1\dots j_p}(x,t)\, dx^{j_1} \wedge \cdots \wedge dx^{j_p},$$

and $x^1,\dots,x^n$ are local coordinates on the manifold $M^n$. We introduce the operator $D$ that associates with a $p$-form $\Omega$ the following $(p-1)$-form $D\Omega$ on $M^n$:

$$D\Omega = \int_0^1 \omega_{p-1}\, dt = \sum_{i_1<\cdots<i_{p-1}} \left(\int_0^1 a_{i_1\dots i_{p-1}}(x,t)\, dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}}.$$

**Lemma 9.4.** *The following identity holds:*

$$dD\Omega - Dd\Omega = (-1)^{p+1}(\Omega|_{t=1} - \Omega|_{t=0}).$$

**Proof.** The statement is obtained by a direct calculation:

$$dD\Omega = \sum_{i_1<\cdots<i_{p-1}} \sum_j \left(\int_0^1 \frac{\partial a_{i_1\dots i_{p-1}}}{\partial x^j}\, dt\right) dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}},$$

$$\begin{aligned}
Dd\Omega &= D(d\omega_p) + D(d\omega_{p-1} \wedge dt)\\
&= D\biggl(\sum_{j_1<\cdots<j_p} \sum_k \frac{\partial b_{j_1\dots j_p}}{\partial x^k}\, dx^k \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_p}\\
&\qquad + \sum_{j_1<\cdots<j_p} \frac{\partial b_{j_1\dots j_p}}{\partial t}\, dt \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_p}\biggr)\\
&\quad + D\biggl(\sum_{i_1<\cdots<i_{p-1}} \sum_j \frac{\partial a_{i_1\dots i_{p-1}}}{\partial x^j}\, dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{p-1}} \wedge dt\biggr)\\
&= (-1)^p \biggl(\sum_{j_1<\cdots<j_p} (b_{j_1\dots j_p}(x,1) - b_{j_1\dots j_p}(x,0))\, dx^{j_1} \wedge \cdots \wedge dx^{j_p}\biggr) + dD\Omega.
\end{aligned}$$

Hence the lemma. □

**Proof of Theorem 9.9.** Let $F\colon M^n \times [0,1]$ be the homotopy between the mappings $f_0$ and $f_1$, and let $\omega$ be a smooth $p$-form on $N^k$. Let $\Omega = F^*\omega$. Lemma 9.4 implies that

$$f_1^*(\omega) - f_0^*(\omega) = (-1)^{p+1}(dDF^*(\omega) - DdF^*(\omega)).$$

But $dF^*(\omega) = F^*(d\omega) = 0$, and we see that

$$[f_1^*(\omega)] - [f_0^*(\omega)] = (-1)^{p+1}[d(DF^*(\omega))] = 0,$$

which implies the equality $f_0^* = f_1^*$ of the homomorphisms of the cohomology rings. □

Two manifolds, $M^n$ and $N^k$, are said to be *homotopically equivalent* if there exist (smooth) mappings $f\colon M^n \to N^k$ and $g\colon N^k \to M^n$ such that the mappings $fg\colon N^k \to N^k$ and $gf\colon M^n \to M^n$ are homotopic to the identity.

EXAMPLE. Euclidean space $\mathbb{R}^n$, the $n$-dimensional disk $\{|x| \leq 1\}$, and a point pt are homotopically equivalent. Indeed, let $f(x) = \mathrm{pt}$ and $g(\mathrm{pt}) = 0$, where $f\colon M^n \to \mathrm{pt}$ and $M^n$ is either $\mathbb{R}^n$ or the disk $\{|x| \leq 1\}$. The mapping $fg$ is the identity, and the mapping $gf$ is homotopic to the identity by means of the homotopy $F(x,t) = (1-t)x$.

It is obvious from the definition of the de Rahm cohomology that a diffeomorphism $f$ gives rise to a cohomology isomorphism. Theorem 9.9 implies a stronger property, namely, the homotopy invariance of the de Rahm cohomology.

**Theorem 9.10.** *For homotopically equivalent manifolds, the corresponding cohomology groups are isomorphic.*

**Proof.** Let $f\colon M^n \to N^k$ and $g\colon N^k \to M^n$ establish the homotopic equivalence. Then by Theorem 9.9 the mappings $f^*g^*\colon H^*(M^k;\mathbb{R}) \to H^*(M^k;\mathbb{R})$ and $g^*f^*\colon H^*(N^n;\mathbb{R}) \to H^*(N^n;\mathbb{R})$ are isomorphisms. Therefore, the mappings $f^*$ and $g^*$ are mutually inverse and establish an isomorphism of the groups $H^*(M^n;\mathbb{R})$ and $H^*(N^k;\mathbb{R})$. Hence the theorem. □

**Corollary 9.4** (Poincaré lemma). *The cohomology groups of the $n$-dimensional disk, the $n$-dimensional Euclidean space $\mathbb{R}^n$, and a point coincide. In particular, any closed $p$-form on the disk or Euclidean space is exact for $p > 0$.*

**9.3.3. Examples of cohomology groups.** By a (geometric) *cycle* on a manifold $M^n$ we will mean a pair $(N^k, f)$ consisting of a closed manifold $N^k$ and a smooth mapping $f\colon N^k \to M^n$. If there is an orientation on the manifold $N^k$, then such a cycle will be said to be oriented.

A cycle $(N^k, f)$ is called a *boundary cycle* or simply a *boundary* if there is a smooth mapping $F\colon P^{k+1} \to M^n$ of a manifold $P^{k+1}$ with boundary $N^k = \partial P^{k+1}$ such that the mapping $F$ coincides on the boundary with the mapping $f$,

$$(N^k, f) = \partial(P^{k+1}, F).$$

In what follows we will assume all the cycles to be oriented, unless otherwise stated.

Any closed $k$-form $\omega$ on a manifold $M^n$ specifies a function on the set of all $k$-dimensional cycles $z = (N^k, f)$ by the formula

$$\langle \omega, z \rangle = \int_{N^k} f^*\omega.$$

**Theorem 9.11.** 1. *The value of the function $\langle \omega, z\rangle$ depends only on the cohomology class of the form $\omega$:*

$$\langle [\omega], z\rangle = \langle \omega + d\varphi, z\rangle = \langle \omega, z\rangle.$$

*In particular, if the form $\omega$ is exact, then*

$$\langle \omega, z\rangle = 0$$

*for all $k$-dimensional cycles.*

2. *The value of the function $\omega(z)$ on a boundary cycle equals zero for any closed form $\omega$:*

$$\langle [\omega], z\rangle = 0 \qquad \text{if} \quad z = \partial(P^{k+1}, F).$$

*In particular, if a cycle $z' = (N^k, f')$ is homotopic to $z$, i.e., the mappings $f$ and $f'$ are smoothly homotopic, then*

$$\langle \omega, z\rangle = \langle \omega, z'\rangle$$

*for any closed form $\omega$.*

**Proof.** Both statements follow from the Stokes formula. We have

$$\begin{aligned}\langle \omega + d\varphi, z\rangle &= \int_{N^k} f^*(\omega + d\varphi) = \int_{N^k} f^*\omega + \int_{N^k} f^* d\varphi \\ &= \int_{N^k} f^*\omega + \int_{N^k} d(f^*\varphi) = \int_{N^k} f^*\omega + \int_{\partial N^k} f^*\varphi = \int_{N^k} f^*\omega = \langle \omega, z\rangle.\end{aligned}$$

This proves the first statement.

For the proof of the second, note that

$$\int_{N^k} f^*\omega = \int_{\partial P^{k+1}} F^*\omega = \int_{P^{k+1}} d(F^*\omega) = \int_{P^{k+1}} F^*(d\omega) = 0.$$

If two cycles are homotopic via the mapping $F\colon N^k \times [0,1] \to M^n$ and $F(x,0) = f(x)$, $F(x,1) = f'(x)$, then the cycle $(\partial(N^k \times [0,1]), F|_{\partial(N^k\times[0,1])})$ is a boundary and

$$\int_{\partial(N^k\times[0,1])} F^*\omega = \int_{N^k} f'^*\omega - \int_{N^k} f^*\omega = \langle[\omega], z'\rangle - \langle[\omega], z\rangle = 0.$$

Hence the theorem. □

**Corollary 9.5.** *If $M^n$ is a closed oriented manifold of dimension $n$, then*

$$H^n(M^n;\mathbb{R}) \neq 0.$$

**Proof.** On a compact manifold one can construct a Riemannian metric $g_{ij}$, and using it, construct the $n$-dimensional volume form $\omega = \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n$. This form is closed, since it has the maximal degree. At the same time, the integral of this form over the $n$-dimensional cycle specified by the manifold $M^n$ itself (with the identity mapping $f\colon M^n \to M^n$) does not vanish,

$$\int_{M^n} \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n > 0.$$

Therefore, $[\omega] \neq 0$ in $H^n(M^n;\mathbb{R})$. Hence the corollary. □

We state the following theorem without proof.

**Theorem 9.12.** *If the integral of a closed $k$-form $\omega$ over any $k$-dimensional cycle is equal to zero, then this form is exact, i.e., $[\omega] = 0$.*

**Corollary 9.6.** *The $p$-dimensional cohomology groups of the $n$-dimensional sphere $S^n$ vanish for $1 \leq p \leq n-1$:*

$$H^p(S^n;\mathbb{R}) = 0, \quad p = 1, \dots, n-1.$$

**Proof.** For $p \leq n-1$ the image of a $p$-dimensional cycle $z$ does not cover the entire $n$-dimensional sphere and hence lies in the complement of some point. This sphere with a point removed is diffeomorphic to the Euclidean space $\mathbb{R}^n$; hence any closed form $\omega$ restricted to the punctured sphere is exact for $p \geq 1$ (see Lemma 9.4). Therefore, the integral of any closed $p$-form $\omega$ over any $p$-dimensional cycle is equal to zero for $1 \leq p \leq n-1$. □

Another way to determine the cohomology groups of the sphere (as well as cohomology of homogeneous spaces) is provided by the averaging method we now present.

Let $M^n$ be a manifold on which a connected compact Lie group $G$ acts,

$$T_g\colon M^n \to M^n, \quad T_{gh} = T_g T_h.$$

If $e$ is the unit element of the group $G$, then the mapping $T_e$ is the identity. Since the group $G$ is connected, for any element $g \in G$ there is a path

$g(t)$ joining it to the unit element, $g(0) = e$, $g(1) = g$. The mapping $T_{g(t)}$ specifies a homotopy between the mapping $T_g$ and the identity mapping $T_e$.

Let $\omega$ be a closed form on $M^n$ and let $T_g^*\omega$ be its image under the mapping $T_g$. Since $T_g$ is homotopic to the identity mapping, Theorem 9.9 implies that

$$\omega = T_g^*\omega - d\psi_g,$$

where $\psi_g$ is a smooth form on $M^n$.

Recall that a form $\omega$ is said to be invariant if $T_g^*\omega = \omega$ for all $g \in G$.

Let $d\mu$ be an invariant volume form on the group $G$ normalized by the condition $\int_G d\mu = 1$.

**Theorem 9.13.** *Let $\omega$ be a closed form. Then the averaging formula*

$$\hat{\omega} = \int_G T_g^*\omega\, d\mu$$

*specifies an invariant form that is closed and homological to the form $\omega$.*

**Proof.** The form $\hat{\omega}$ is invariant by construction, being the integral of all possible translations of the form $\omega$.

We have

$$\hat{\omega} = \int_G (\omega + d\psi_g)\, d\mu = \omega \int_G d\mu + d\Big(\int_G \psi_g\, d\mu\Big) = \omega + d\Big(\int_G \psi_g\, d\mu\Big).$$

Therefore, the forms $\omega$ and $\hat{\omega}$ are cohomological. □

**Corollary 9.7.** *The cohomology ring of the $n$-dimensional torus is generated by one-dimensional classes $\tau_1, \dots, \tau_n$ satisfying only the relations $\tau_i\tau_j = -\tau_j\tau_i$ for $i, j = 1, \dots, n$. Therefore*

$$\dim H^p(T^n; \mathbb{R}) = \frac{n!}{p!\,(n-p)!}.$$

**Proof.** We represent the torus as the direct product of $n$ circles with coordinates $\varphi_1, \dots, \varphi_n$ defined modulo $2\pi$. The torus acts on itself by translations

$$(\varphi_1, \dots, \varphi_n) \to (\varphi_1 + t_1, \dots, \varphi_n + t_n).$$

Obviously, the invariant forms are precisely the forms with constant coefficients, hence any invariant form is closed.

The forms $d\varphi_{i_1} \wedge \dots \wedge d\varphi_{i_k}$ with $i_1 < \dots < i_k$ constitute a basis in the space of invariant $k$-forms. Let us define a $k$-dimensional cycle $z_{j_1 \dots j_k}$, where $j_1 < \dots < j_k$, by the equations $\varphi_i = 0$ for $i \neq j_1, \dots, j_k$. We have

$$\int_{z_{j_1 \dots j_k}} d\varphi_{i_1} \wedge \dots \wedge d\varphi_{i_k} = \begin{cases} (2\pi)^k & \text{for } (i_1, \dots, i_k) = (j_1, \dots, j_k), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, all the forms $d\varphi_{i_1} \wedge \cdots \wedge d\varphi_{i_k}$ are pairwise noncohomological and specify nontrivial cohomology classes. It remains to note that the cohomology space as a ring is generated by the classes $\tau_1 = [d\varphi_1], \dots, \tau_n = [d\varphi_n]$. This proves the corollary. □

For $n = 1$ we obtain the cohomology group of the circle

$$H^0(S^1;\mathbb{R}) = H^1(S^1;\mathbb{R}) = \mathbb{R}, \quad H^k(S^1;\mathbb{R}) = 0 \quad \text{for} \quad k > 1,$$

and conclude that a periodic 1-form $\omega = f(\varphi)\,d\varphi$ on the circle specifies a nontrivial cohomology class if and only if $\int_0^{2\pi} f(\varphi)\,d\varphi \neq 0$.

We have shown in 6.1.5 that the $n$-dimensional sphere is a homogeneous space of the form

$$S^n = \mathrm{SO}(n+1)/\mathrm{SO}(n).$$

Any closed form is cohomological to an invariant form, and there is a unique, up to a constant factor, $n$-dimensional form on the sphere $S^n$ (which is uniquely determined by its value $C\,dx^1 \wedge \cdots \wedge dx^n$ at an arbitrary point). Therefore $H^n(S^n;\mathbb{R}) = \mathbb{R}$.

**Lemma 9.5.** *In the space $\mathbb{R}^n$ there are no nonzero skew-symmetric tensors of rank $p < n$ that are invariant with respect to the group* $\mathrm{SO}(n)$.

**Proof.** Let $T$ be a tensor of rank $k$ that is invariant with respect to the group $\mathrm{SO}(n)$. Let $x$ and $y$ be a pair of coordinates from among the orthogonal coordinates $x^1, \dots, x^n$. The tensor $T$ can be written as

$$T = dx \wedge a + dy \wedge b + dx \wedge dy \wedge c,$$

where the forms $a, b$, and $c$ written in the orthogonal coordinates do not contain $dx$ and $dy$. Consider the rotation

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

It acts on the forms as follows:

$$dx \to \cos\varphi\,dx - \sin\varphi\,dy, \quad dy \to \sin\varphi\,dx + \cos\varphi\,dy, \quad a \to a,\ b \to b,\ c \to c,$$

and consequently,

$$T \to dx \wedge (\cos\varphi a + \sin\varphi b) + dy \wedge (-\sin\varphi a + \cos\varphi b) + dx \wedge dy \wedge c.$$

Hence we see that the tensor $T$ is invariant under rotations through any angle $\varphi$ if and only if $a = b = 0$, i.e., $T = dx \wedge dy \wedge c$. Since the pair of orthogonal coordinates was chosen arbitrarily, the invariant tensor $T$ has the form $f(x)\,dx^1 \wedge \cdots \wedge dx^n$, i.e., has the maximal rank $n$ and is proportional to the volume form. Hence the lemma. □

**Corollary 9.8.** *There are no invariant forms of rank $p = 1, \dots, n-1$ on the sphere $S^n = \mathrm{SO}(n+1)/\mathrm{SO}(n)$.*

**Proof.** Each invariant form is specified by its value at a single point of the sphere, and it must be invariant with respect to the action of the stationary subgroup $\mathrm{SO}(n)$. But we know that on $\mathbb{R}^n$ there are no skew-symmetric tensors of rank $p = 1, \dots, n-1$. Hence the statement. □

**Corollary 9.9.** *The cohomology groups of the $n$-dimensional sphere $S^n$ are*

$$H^p(S^n; \mathbb{R}) = \begin{cases} \mathbb{R} & \textit{for } p = 0 \textit{ and } n, \\ 0 & \textit{otherwise.} \end{cases}$$

Now we state, for reference, some useful facts.

It is not true that if the integral of any closed $p$-form over a $p$-dimensional cycle $(N^p, f)$ vanishes, then this cycle is a boundary. For example, if $N^p$ is a closed manifold which is not a boundary of any other manifold (and such manifolds do exist) and $f: N^p \to M^n$ maps it into a point, then the integral of any form over this cycle equals zero, but it is not a boundary cycle.

The quotient space of all oriented $p$-cycles by the boundary cycles is called the *group of oriented bordisms* $\Omega_p^{\mathrm{SO}}(M^n)$. It is nontrivial already for points. If we consider also the mappings of nonoriented manifolds $N^p$, then the quotient group of such cycles by the boundary cycles is called the *group of bordisms* $\Omega_p^{\mathrm{O}}(M^n)$.

Let $C_p(M^n)$ be the space generated by all oriented $p$-cycles over the field $\mathbb{R}$. For each element $z$ of this space we can define the integral of a closed $p$-form $\omega$ over $z$. Namely, let $z = \lambda_1 z_1 + \dots + \lambda_k z_k$, where $\lambda_i \in \mathbb{R}$ and the $z_i = (N_i^p, f_i)$ are ordinary cycles, $i = 1, \dots, k$. Then

$$\langle \omega, z \rangle = \sum_i \lambda_i \langle \omega, z_i \rangle = \sum_i \lambda_i \int_{N_i} f_i^* \omega.$$

Let $B_p(M^n) \subset C_p(M^n)$ be the subspace consisting of the cycles such that the integrals of all closed $p$-forms over these cycles vanish. The quotient space

$$H_p(M^n; \mathbb{R}) = C_p(M^n)/B_p(M^n)$$

is called the real $p$-dimensional *homology* group of the manifold $M^n$. Since all the integrals over boundary cycles vanish, we obtain a homomorphism

$$\Omega_p^{\mathrm{SO}}(M^n) \otimes \mathbb{R} \to H_p(M^n; \mathbb{R}).$$

A cycle lying in the kernel of this homomorphism is said to be *homological to zero*.

The image of this homomorphism coincides with the entire homology group. This is a deep result of algebraic topology.

The homology space is dual to the cohomology group $H^p(M^n;\mathbb{R})$, and for any nonzero element $[\omega] \in H^p(M^n;\mathbb{R})$ the linear functional

$$\langle [\omega],[z]\rangle = \langle \omega, z\rangle, \quad [z] \in H_p(M^n;\mathbb{R}),$$

is nondegenerate. Furthermore, any smooth mapping

$$g\colon M^n \to V^k$$

induces homomorphisms (for all $p$) of cohomology and homology groups:

$$g^*\colon H^p(V^k;\mathbb{R}) \to H^p(M^n;\mathbb{R}), \quad g^*[\omega] = [g^*\omega],$$
$$g_*\colon H_p(M^n;\mathbb{R}) \to H_p(V^k;\mathbb{R}), \quad g_*[(N,f)] = [(N,gf)],$$

and by construction these homomorphisms are dual to each other,

$$\langle g^*[\omega],[z]\rangle = \langle [\omega], g_*[z]\rangle.$$

For compact manifolds all the cohomology groups are finite-dimensional and their dimensions are called the *Betti numbers*:

$$b_p(M^n) = \dim H^p(M^n;\mathbb{R}) = \dim H_p(M^n;\mathbb{R}).$$

The alternating sum of the Betti numbers,

$$\chi(M^n) = b_0(M^n) - b_1(M^n) + b_2(M^n) - \cdots + (-1)^k b_k(M^n) + \cdots,$$

is called the *Euler characteristic* of the manifold $M^n$.

Given an $n$-dimensional oriented connected compact manifold $M^n$ without boundary, its cohomology groups meet the Poincaré duality:

$$H^{n-k}(M^n;\mathbb{R}) = H^k(M^n;\mathbb{R}), \quad k = 0,1,\ldots,n$$

(see Problem 14.10), and therefore, for such an odd-dimensional manifold, the Euler characteristic vanishes: $\chi(M^{2n+1}) = 0$.

## Exercises to Chapter 9

1. Prove that the mapping $F^*$ induced by a smooth mapping $F$ commutes with the exterior differential,

$$F^*(d\omega) = dF^*(\omega).$$

Using this fact, prove that the exterior differential commutes with the Lie derivative,

$$L_X(d\omega) = d(L_X\omega).$$

2. Let us associate with each vector field $X$ the following linear operator $i(X)$ on the forms:

$$[i(X)\omega]_{j_1\ldots j_{k-1}} = X^j \omega_{jj_1\ldots j_{k-1}}.$$

Prove that

a) $i(X)$ is antidifferentiation:

$$i(X)(\omega_1 \wedge \omega_2) = (i(X)\omega_1) \wedge \omega_2 + (-1)^k \omega_1 \wedge i(X)\omega_2,$$

where $\omega_1$ is a $k$-form;

b) the following formula holds:

$$i(X)\, d + d\, i(X) = L_X,$$

where $L_X$ is the Lie derivative in the direction of the field $X$.

3. Let $X$ be a vector field in $\mathbb{R}^4$. The flux of the field $X$ over a hypersurface is the integral of the differential form $\omega = X^i\, dS_i$, where

$$dS_i = \frac{1}{6} \sqrt{|g|}\, \varepsilon_{jkli}\, dx^j \wedge dx^k \wedge dx^l.$$

Prove that if a hypersurface $\Gamma$ is the boundary of a domain $U$, then

$$\int_{\partial U = \Gamma} X^i\, dS_i = \int_U \frac{\partial X^i}{\partial x^i}\, d^4 x.$$

4. Let the charge density $\rho$ be equal to zero, and let the vector potential of the electromagnetic field $A_i$ have the form $A_i(x - ct)$, $i = 0, 1, 2, 3$ (electromagnetic wave propagating along the $x$-axis). Prove that all the eigenvalues of the electromagnetic field are identically equal to zero.

5. Prove that the Berezinian and supertrace satisfy the relation

$$\operatorname{Ber}(1 + \varepsilon A) = 1 + \varepsilon \operatorname{str} A + O(\varepsilon^2).$$

6. Determine the value of the Gaussian integral over the superspace:

$$\int_{\mathbb{R}^{m|n}} \exp(a_{ij} x^i x^j + 2 b_{ij} x^i \xi_j + c_{ij} \xi_i \xi_j)\, d^m x \wedge d^n \xi.$$

Here the summation rule over repeated indices is applied, the quantities $a_{ij}$ and $c_{ij}$ are even and $b_{ij}$ are odd, and the following conditions are satisfied: $a_{ij} = a_{ji}$, $b_{ij} = -b_{ji}$, $c_{ij} = -c_{ji}$.

7. Prove that the cohomology groups of the space $\mathbb{C}P^n$ are

$$H^p(\mathbb{C}P^n; \mathbb{R}) = \begin{cases} \mathbb{R} & \text{for } p = 2k \leq 2n, \\ 0 & \text{otherwise}, \end{cases}$$

and the cohomology ring is multiplicatively generated by the cohomology class of the Kähler form $[\omega]$ satisfying the only condition $[\omega]^{n+1} = 0$.

8. Prove that the cohomology groups of the direct product of manifolds $M$ and $N$ are given by

$$H^p(M \times N; \mathbb{R}) = \sum_{k+l=p} H^k(M; \mathbb{R}) \otimes H^l(N; \mathbb{R}).$$

Chapter 10

# Connections and Curvature

## 10.1. Covariant differentiation

**10.1.1. Covariant differentiation of vector fields.** In order to construct differential calculus for tensors, we need to define the gradient of a tensor so that the resulting object will again be a tensor. The usual definition of gradient does not meet this requirement because, in general, the gradient transforms as a tensor only under affine transformations.

**Theorem 10.1.** *Let $T^{i_1\dots i_p}_{j_1\dots j_q}$ be a tensor field of type $(p,q)$ in the space with coordinates $x^1,\dots,x^n$, and let*

$$z^i = a^i_j x^j + b^i, \quad i,j = 1,\dots,n,$$

*where $a^i_j = \text{const}$, be an affine transformation specifying the transition to new coordinates $z^1,\dots,z^n$. Then under this change of coordinates the gradient*

$$T^{i_1\dots i_p}_{j_1\dots j_q,k} = \frac{\partial T^{i_1\dots i_p}_{j_1\dots j_q}}{\partial x^k} \tag{10.1}$$

*transforms as a tensor of type $(p,q+1)$.*

**Proof.** For brevity, we restrict ourselves to the case of a vector field because the general case is treated similarly. The equality

$$\tilde{T}^i = T^j \frac{\partial z^i}{\partial x^j}$$

implies that

$$(10.2)\qquad \tilde{T}^i_{,k} = \frac{\partial \tilde{T}^i}{\partial z^k} = \frac{\partial x^l}{\partial z^k}\frac{\partial}{\partial x^l}\left(T^j\frac{\partial z^i}{\partial x^j}\right) = \frac{\partial x^l}{\partial z^k}\frac{\partial z^i}{\partial x^j}T^j_{,l} + \frac{\partial x^l}{\partial z^k}T^j\frac{\partial^2 z^i}{\partial x^j \partial x^l}.$$

A transformation is affine if and only if

$$(10.3)\qquad \frac{\partial^2 z^i}{\partial x^j \partial x^l} = 0, \quad i, j, l = 1, \dots, n,$$

at each point. This equality implies that affine changes of coordinates satisfy the equality

$$\tilde{T}^i_{,k} = \frac{\partial x^l}{\partial z^k}\frac{\partial z^i}{\partial x^j}T^j_{,l}.$$

In a similar way one proves that the gradient of any tensor field transforms as a tensor provided equality (10.3) holds everywhere. □

Given any nonaffine change of coordinates, one can always find a vector field $T^i$ such that the second term in formula (10.2) does not vanish. This means that the gradient of the field $T$ is not a tensor.

Hence we conclude that the operation of taking gradient must be generalized to account for the geometry of the space, and in the case of Euclidean geometry, this operation should coincide with the usual gradient (10.1). The result of this operation must be a tensor.

Consider two examples which will lead us to the required formula for such an extension.

EXAMPLES. 1. EUCLIDEAN CONNECTION. Suppose that a generalized gradient operation $\nabla$ is defined on vector fields in such a way that $\nabla$ maps vector fields into tensors of type $(1,1)$:

$$(10.4)\qquad \tilde{T}^i_{;j} = T^p_{;q}\frac{\partial z^i}{\partial x^p}\frac{\partial x^q}{\partial z^j}.$$

Moreover, suppose that in some coordinate system $(x^1, \dots, x^n)$ this operation is represented as the usual gradient,

$$\nabla_j T^i = T^i_{,j} = \frac{\partial T^i}{\partial x^j}.$$

Let $(z^1, \dots, z^n)$ be another system of coordinates. How is the operation $\nabla$ written in this coordinate system?

Since $T^p_{;q} = \frac{\partial T^p}{\partial x^q}$, formula (10.4) becomes

$$\tilde{T}^i_{;j} = \frac{\partial T^p}{\partial x^q}\frac{\partial x^q}{\partial z^j}\frac{\partial z^i}{\partial x^p} = \frac{\partial T^p}{\partial z^j}\frac{\partial z^i}{\partial x^p}.$$

Since $\tilde{T}^i = T^p \frac{\partial z^i}{\partial x^p}$, this equality takes the form

$$\tilde{T}^i_{;j} = \frac{\partial T^p}{\partial z^j}\frac{\partial z^i}{\partial x^p} = \frac{\partial \tilde{T}^i}{\partial z^j} - T^p \frac{\partial}{\partial z^j}\Big(\frac{\partial z^i}{\partial x^p}\Big) = \frac{\partial \tilde{T}^i}{\partial z^j} - T^p \frac{\partial^2 z^i}{\partial x^p \partial x^q}\frac{\partial x^q}{\partial z^j}.$$

Substituting $T^p = \tilde{T}^k \frac{\partial x^p}{\partial z^k}$ into this relation, we obtain

$$\tilde{T}^i_{;j} = \frac{\partial \tilde{T}^i}{\partial z^j} - \tilde{T}^k \frac{\partial x^p}{\partial z^k}\frac{\partial^2 z^i}{\partial x^p \partial x^q}\frac{\partial x^q}{\partial z^j}.$$

Using the notation

$$\Gamma^i_{kj} = -\frac{\partial x^p}{\partial z^k}\frac{\partial x^q}{\partial z^j}\frac{\partial^2 z^i}{\partial x^p \partial x^q}, \tag{10.5}$$

we arrive at the equality

$$\tilde{T}^i_{;j} = \frac{\partial \tilde{T}^i}{\partial z^j} + \Gamma^i_{kj}\tilde{T}^k. \tag{10.6}$$

Thus we come to the following conclusion.

**Theorem 10.2.** *Suppose that the "generalized gradient" of a vector field $(T^i)$ transforms as a tensor under any change of coordinates, and in coordinates $(x)$ it is calculated by the usual formula*

$$T^i_{;j} = \frac{\partial T^i}{\partial x^j}.$$

*Then in any other coordinate system $(z)$ this gradient has the form*

$$\tilde{T}^i_{;j} = \frac{\partial \tilde{T}^i}{\partial z^j} + \Gamma^i_{kj}\tilde{T}^k,$$

*where the coefficients $\Gamma^i_{sj}$ are given by formula* (10.5).

2. Covariant differentiation on a surface. Let $r(x^1, x^2)$ be a surface imbedded into three-dimensional Euclidean space $\mathbb{R}^3$ with basis $e_1, e_2, e_3$. Denote by $r_1 = \partial r/\partial x^1$ and $r_2 = \partial r/\partial x^2$ the basis vectors in the tangent spaces, and by $n$ the field of unit normal vectors to the surface.

Recall the derivational equations

$$r_{ij} = \frac{\partial^2 r}{\partial x^i \partial x^j} = \Gamma^k_{ij} r_k + b_{ij} n.$$

Suppose we are given a vector field $T = T^i r_i$ on the surface. We differentiate it at some point in the direction of the vector $v = v^j r_j$:

$$\partial_v T = v^j \frac{\partial T^i}{\partial x^j} r_i + v^j T^i \frac{\partial r_i}{\partial x^j} = v^j \frac{\partial T^i}{\partial x^j} r_i + v^j T^i (\Gamma^k_{ij} r_k + b_{ij} n).$$

For an "observer" living on the surface, the vector $n$ is unobservable because it is not a tangent vector. Trying to build the intrinsic geometry of the surface, we must forget about this vector, since it does not exist for such an

"observer". Hence we define the "generalized derivative" $\nabla$ of the field $T$ in the direction $v$ as the projection of $\partial_v T$ to the tangent space to the surface:

$$\nabla_v T = v^j \frac{\partial T^i}{\partial x^j} r_i + v^j T^i \Gamma^k_{ij} r_k = T^i_{;j} v^j r_i.$$

In particular,

$$\nabla_{r_j} r_i = \Gamma^k_{ij} r_k. \tag{10.7}$$

Thus we arrive at the following definition of the generalized gradient of a vector field on a surface:

$$T^i_{;j} = \nabla_{r_j} T^i = \frac{\partial T^i}{\partial x^j} + \Gamma^i_{kj} T^k. \tag{10.8}$$

This quantity is a tensor since its definition is invariant and does not depend on coordinates. Indeed, the field $T$ can be expanded in the basis of $\mathbb{R}^3$:

$$T = \hat{T}^1 e_1 + \hat{T}^2 e_2 + \hat{T}^3 e_3,$$

where the $\hat{T}^i$ are functions of $x^1, x^2$, $i = 1, 2, 3$. Then $\partial_v T = \frac{\partial \hat{T}^i}{\partial x^j} v^j e_i$ and this vector as well as its projection to the tangent space do not depend on the choice of coordinates on the surface. Therefore, $T^i_{;j}$ is a tensor field of type $(1,1)$ on the surface.

If the surface is a Euclidean plane in $\mathbb{R}^3$ with linear coordinates $x^1$ and $x^2$, then all the $\Gamma^i_{jk}$ vanish and the generalized gradient coincides with the ordinary gradient.

**Theorem 10.3.** *The "generalized derivative" $\nabla_v T$ of a vector field $T$ on a surface in $\mathbb{R}^3$ defined as the projection of $\partial_v T$ to the tangent space to the surface by formula* (10.8) *is linear in $v$ and $T$,*

$$\nabla_v T = T^i_{;j} v^j,$$

*and $T^i_{;j}$ is a tensor of type $(1,1)$.*

Formula (10.6) is the same as (10.8). We will take this formula for the definition of the "generalized gradient" in the general case. Now we will find conditions on the symbols $\Gamma^i_{jk}$ which ensure that the gradient $T^i_{;j}$ is a tensor.

**Theorem 10.4.** *Suppose that the "generalized gradient" $T^i_{;j}$ of a vector field $T$ is a tensor that, in coordinate systems $(x)$ and $(z)$, is given by the rules*

$$T^i_{;j} = \frac{\partial T^i}{\partial x^j} + \Gamma^i_{kj} T^k \quad \textit{and} \quad \tilde{T}^p_{;q} = \frac{\partial \tilde{T}^p}{\partial z^q} + \tilde{\Gamma}^p_{sq} \tilde{T}^q.$$

*Then the symbols $\Gamma^i_{kj}$ and $\tilde{\Gamma}^p_{sq}$ are related by the equation*

$$\Gamma^i_{kj} = \tilde{\Gamma}^p_{sq} \frac{\partial x^i}{\partial z^p} \frac{\partial z^q}{\partial x^j} \frac{\partial z^s}{\partial x^k} + \frac{\partial x^i}{\partial z^p} \frac{\partial^2 z^p}{\partial x^k \partial x^j}. \tag{10.9}$$

**Proof.** Since $T^i$ and $T^i_{;j}$ are tensors, they satisfy the equalities

$$T^i_{;j} = \frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\tilde{T}^p_{;q}, \quad \tilde{T}^p = T^k\frac{\partial z^p}{\partial x^k}, \quad \tilde{T}^s = T^k\frac{\partial z^s}{\partial x^k},$$

and we have

$$\begin{aligned}
T^i_{;j} &= \frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\tilde{T}^p_{;q} = \frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\Big(\frac{\partial \tilde{T}^p}{\partial z^q} + \tilde{\Gamma}^p_{sq}\tilde{T}^s\Big) \\
&= \frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\Big(\frac{\partial}{\partial z^q}\Big(T^k\frac{\partial z^p}{\partial x^k}\Big) + \tilde{\Gamma}^p_{sq}T^k\frac{\partial z^s}{\partial x^k}\Big) \\
&= \frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\Big(\frac{\partial T^k}{\partial z^q}\frac{\partial z^p}{\partial x^k} + T^k\frac{\partial}{\partial z^q}\Big(\frac{\partial z^p}{\partial x^k}\Big) + \tilde{\Gamma}^p_{sq}T^k\frac{\partial z^s}{\partial x^k}\Big) \\
&= \Big(\frac{\partial x^i}{\partial z^p}\frac{\partial z^p}{\partial x^k}\Big)\frac{\partial T^k}{\partial z^q}\frac{\partial z^q}{\partial x^j} \\
&\quad + \Big(\frac{\partial x^i}{\partial z^p}\Big(\frac{\partial z^q}{\partial x^j}\frac{\partial x^m}{\partial z^q}\Big)\frac{\partial^2 z^p}{\partial x^k\partial x^m} + \tilde{\Gamma}^p_{sq}\frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\frac{\partial z^s}{\partial x^k}\Big)T^k.
\end{aligned}$$

Since $\frac{\partial x^i}{\partial z^p}\frac{\partial z^p}{\partial x^k} = \delta^i_k$ and $\frac{\partial z^q}{\partial x^j}\frac{\partial x^m}{\partial z^q} = \delta^m_j$, the last formula becomes

$$T^i_{;j} = \frac{\partial T^i}{\partial z^q}\frac{\partial z^q}{\partial x^j} + \Big(\frac{\partial x^i}{\partial z^p}\frac{\partial^2 z^p}{\partial x^k\partial x^j} + \tilde{\Gamma}^p_{sq}\frac{\partial x^i}{\partial z^p}\frac{\partial z^q}{\partial x^j}\frac{\partial z^s}{\partial x^k}\Big)T^k.$$

Comparing this expression with the formula

$$T^i_{;j} = \frac{\partial T^i}{\partial x^j} + \Gamma^i_{kj}T^k$$

and taking into account that $T$ is an arbitrary vector field, we conclude that the symbols $\Gamma^i_{kj}$ and $\tilde{\Gamma}^p_{sq}$ must satisfy relation (10.9). □

Now we are prepared to define the generalized gradient of vector fields.

We say that the operation of *covariant differentiation* (taking the gradient) is specified if in any system of coordinates $x^1, \dots, x^n$ there is an array of functions $\Gamma^i_{kj}(x)$ that transforms under coordinate changes $z = z(x)$ by the formula

$$\Gamma^i_{kj} \to \tilde{\Gamma}^p_{sq} = \Gamma^i_{kj}\frac{\partial z^p}{\partial x^i}\frac{\partial x^j}{\partial z^q}\frac{\partial x^k}{\partial z^s} + \frac{\partial z^p}{\partial x^i}\frac{\partial^2 x^i}{\partial z^s\partial z^q}. \tag{10.10}$$

The *covariant derivative* of a vector field $(T^i)$ is the following tensor of type $(1,1)$:

$$\nabla_j T^i = T^i_{;j} = \frac{\partial T^i}{\partial x^j} + \Gamma^i_{kj}T^k. \tag{10.11}$$

The quantities $\Gamma^i_{kj}$ are called the *Christoffel symbols*.

We denote the covariant gradient by $T^i_{;j}$ to distinguish it from the Euclidean gradient $T^i_{,j} = \frac{\partial T^i}{\partial x^j}$:

$$\nabla_j T^i = T^i_{;j}, \quad \frac{\partial T^i}{\partial x^j} = T^i_{,j}.$$

The operation of covariant differentiation is also referred to as a *linear connection* on tangent vector fields. A linear connection is said to be *Euclidean* if there are coordinates $x^1, \dots, x^n$ in which $\Gamma^k_{ij} = 0$, or, equivalently,

$$T^i_{;j} = \frac{\partial T^i}{\partial x^j}.$$

Such coordinates are called *Euclidean.*

In the general case, take some basis of vector fields $e_1, \dots, e_n$ and consider their covariant derivatives. Relation (10.11) implies the following formula, which explains the meaning of the Christoffel symbols:

$$\nabla_j e_i = \Gamma^k_{ij} e_k.$$

For covariant differentiation on surfaces in $\mathbb{R}^3$ this formula turns into relation (10.7).

Before defining the covariant differentiation of an arbitrary tensor field, we point out some properties of Christoffel symbols. It is seen directly from (10.10) that due to the presence of the second term $\frac{\partial z^p}{\partial x^i} \frac{\partial^2 x^i}{\partial z^s \partial z^q}$ the Christoffel symbols do not form a tensor. However, under affine transformations of coordinates this term vanishes. It also vanishes in the transformation formulas for the difference of the Christoffel symbols of two connections and for the alternating Christoffel symbol (because of its symmetry in $q$ and $s$). Thus we have proved the following lemma.

**Lemma 10.1.** 1. *The Christoffel symbols $\Gamma^i_{kj}$ transform as a tensor only under linear or affine transformations of coordinates $z^i = z^i(x^1, \dots, x^n)$. In this case $\frac{\partial^2 z^i}{\partial x^k \partial x^j} \equiv 0$ for all $i, k, j$.*

2. *The variation (difference) of Christoffel symbols of two different connections*

$$\delta\Gamma^i_{jk} = \Gamma'^i_{jk} - \Gamma^i_{jk}$$

*is a tensor of type $(1, 2)$.*

3. *The alternated expression*

$$T^i_{kj} = \Gamma^i_{kj} - \Gamma^i_{jk} = \Gamma^i_{[kj]} \tag{10.12}$$

*forms a tensor, which is called the torsion tensor.*

A connection $\Gamma^i_{kj}$ is said to be *symmetric* if its *torsion tensor* $T^i_{kj} = \Gamma^i_{[kj]}$ is identically equal to zero, $\Gamma^i_{kj} = \Gamma^i_{jk}$.

For example, for a Euclidean connection, $\Gamma^i_{kj} \equiv 0$ in Euclidean coordinates. In these coordinates $T^i_{jk} \equiv 0$, and, since the $T^i_{jk}$ form a tensor, they vanish in any coordinates. Therefore, a Euclidean connection is symmetric.

A symmetric connection has the following property.

**Lemma 10.2.** *If a connection is symmetric, then in a neighborhood of any point $x_0 \in M^n$ there are coordinates $(z^1, \ldots, z^n)$ in which the Christoffel symbols at this point vanish,*

$$\Gamma^i_{jk}(x_0) = 0.$$

**Proof.** Recall the transformation rule for Christoffel symbols $\Gamma^i_{kj} \to \tilde{\Gamma}^p_{sq}$ under transitions from coordinates $(x)$ to coordinates $(z)$:

$$\Gamma^i_{kj} \to \tilde{\Gamma}^p_{sq} = \Gamma^i_{kj} \frac{\partial z^p}{\partial x^i} \frac{\partial x^j}{\partial z^q} \frac{\partial x^k}{\partial z^s} + \frac{\partial z^p}{\partial x^i} \frac{\partial^2 x^i}{\partial z^s \partial z^q}.$$

We must find a change of coordinates such that $\tilde{\Gamma}^p_{sq}(x_0) = 0$, i.e., such that at the point $x_0 \in M^n$ the following relations hold:

$$\Gamma^i_{kj} \frac{\partial z^p}{\partial x^i} \frac{\partial x^j}{\partial z^q} \frac{\partial x^k}{\partial z^s} + \frac{\partial z^p}{\partial x^i} \frac{\partial^2 x^i}{\partial z^s \partial z^q} = 0. \tag{10.13}$$

We may assume that the coordinates of the point $x_0$ are equal to zero in both coordinate systems. Define the functions $z^1, \ldots, z^n$ in a neighborhood of zero by the formulas

$$x^i = z^i - \Gamma^i_{jk} z^j z^k.$$

By the implicit function theorem this mapping is invertible in a neighborhood of zero, since its Jacobian at zero is the identity matrix. Obviously, relation (10.13) is fulfilled under this change of coordinates, hence $(z^1, \ldots, z^n)$ are the required coordinates. □

**10.1.2. Covariant differentiation of tensors.** We will extend covariant differentiation to arbitrary tensors with the aid of the following rules:

1) The operation of covariant differentiation is linear,

$$\nabla_i(T + S) = \nabla_i T + \nabla_i S$$

for any tensor fields $T$ and $S$.

2) The covariant derivative of a scalar tensor field (a function) coincides with the ordinary gradient,

$$\nabla_i f = f_{;i} = \frac{\partial f}{\partial x^i}. \tag{10.14}$$

3) The covariant derivative of the contraction of tensors $T^{(k)}_{m(j)} S^{m(p)}_{(q)}$ with respect to the index $m$ is calculated by the Leibniz formula,

$$\nabla_i (T^{(k)}_{m(j)} S^{m(p)}_{(q)}) = (\nabla_i T^{(k)}_{m(j)}) S^{m(p)}_{(q)} + T^{(k)}_{m(j)} (\nabla_i S^{m(p)}_{(q)}). \tag{10.15}$$

4) The covariant derivative of a vector field is given by (10.11).

These rules uniquely define covariant differentiation for all tensors. Indeed, we first derive the formula for the covariant derivative of a covector field.

Let $T_i$ be a covector field. For any vector field $S$ the contraction $T_iS^i$ is a scalar, hence formula (10.14) implies that

$$\nabla_j(T_iS^i) = \frac{\partial(T_iS^i)}{\partial x^j} = \frac{\partial T_i}{\partial x^j}S^i + T_i\frac{\partial S^i}{\partial x^j}.$$

The Leibniz formula (see (10.15)) implies that

$$\nabla_j(T_iS^i) = T_{i;j}S^i + T_iS^i_{;j} = T_{i;j}S^i + T_i\Big(\frac{\partial S^i}{\partial x^j} + \Gamma^i_{kj}S^k\Big).$$

Comparing these two expressions for $\nabla_j(T_iS^i)$ we obtain

$$\Big(T_{i;j} - \frac{\partial T_i}{\partial x^j} + \Gamma^k_{ij}T_k\Big)S^i = 0.$$

Since $S$ is an arbitrary vector, the expression in parentheses vanishes. Thus we have proved the following fact.

**Theorem 10.5.** *The covariant derivative of a vector field is given by the formula*

$$\nabla_jT_i = T_{i;j} = \frac{\partial T_i}{\partial x^j} - \Gamma^k_{ij}T_k. \tag{10.16}$$

Now we can derive the general formula by induction. Assume that it has been derived for tensors of rank at most $d$ and let $T$ be a tensor of type $(m+1,n)$ or $(m,n+1)$, where $m+n=d$. Assume without loss of generality that the type is $(m,n+1)$. We already know the formulas for $\nabla_j(T^{(p)}_{i(q)}S^i)$ and for $\nabla_jS^i$. Using the Leibniz formula (10.15) in the same manner as when deriving (10.16), we obtain the formula for $\nabla_jT$.

We omit the cumbersome calculations and state only the final result.

**Theorem 10.6.** *The covariant derivative of a tensor $T^{(i)}_{(k)}$ of type $(m,n)$, where $(i)=i_1,\dots,i_m$, $(k)=k_1,\dots,k_n$, is*

$$\begin{aligned}\nabla_jT^{(i)}_{(k)} = \frac{\partial T^{(i)}_{(k)}}{\partial x^j} - T^{(i)}_{kk_2\dots k_n}\Gamma^k_{k_1j} - \dots - T^{(i)}_{k_1\dots k}\Gamma^k_{k_nj}\\ + T^{ii_2\dots i_m}_{(k)}\Gamma^{i_1}_{ij} + \dots + T^{i_1\dots i}_{(k)}\Gamma^{i_m}_{ij}.\end{aligned} \tag{10.17}$$

*In particular, the covariant derivatives of tensors of rank two are*

$$\nabla_j T^{ik} = \frac{\partial T^{ik}}{\partial x^j} + \Gamma^i_{lj} T^{lk} + \Gamma^k_{lj} T^{il},$$
$$\nabla_j T^i_k = \frac{\partial T^i_k}{\partial x^j} + \Gamma^i_{lj} T^l_k - \Gamma^l_{kj} T^i_l,$$
$$\nabla_j T_{ik} = \frac{\partial T_{ik}}{\partial x^j} - \Gamma^l_{ij} T_{lk} - \Gamma^l_{kj} T_{il}.$$

Now we will state two corollaries from formula (10.17) for the covariant derivative. Their proofs are left as an exercise for the reader.

**Corollary 10.1.** *Covariant differentiation satisfies the Leibniz rule*

$$\nabla_j (R^{(i)}_{(k)} \otimes S^{(p)}_{(q)}) = (\nabla_j R^{(i)}_{(k)}) \otimes S^{(p)}_{(q)} + R^{(i)}_{(k)} \otimes (\nabla_j S^{(p)}_{(q)}),$$

*which expresses the derivative of the tensor product in terms of the derivatives of the factors.*

Assume now that the tensor field $T^{(i)}_{(k)}$ is defined only on a curve $x(t)$, and let $v = \dot{x}$ be the velocity vector of the curve. We can define the *covariant derivative along the curve*:

$$\begin{aligned}\nabla_v T = \frac{dT}{dt} - T^{(i)}_{kk_2\ldots k_q} \Gamma^k_{k_1 j} v^j - \cdots - T^{(i)}_{k_1\ldots k_{q-1}k} \Gamma^k_{k_q j} v^j \\ + T^{ii_2\ldots i_p}_{(k)} \Gamma^{i_1}_{ij} v^j + \cdots + T^{i_1\ldots i_{p-1} i}_{(k)} \Gamma^{i_p}_{ij} v^j.\end{aligned} \tag{10.18}$$

**Corollary 10.2.** *The covariant derivative $\nabla_v T$ of the field $T$ along the curve is well defined, and its value depends only on the values of $T$ on this curve.*

### 10.1.3. Gauge fields.

Let $G \subset \mathrm{GL}(m, \mathbb{R})$ be a matrix group acting on $\mathbb{R}^k$.

Let $D$ be an $n$-dimensional domain with coordinates $x^1, \ldots, x^n$. A vector-function in $D$ is simply a function

$$\psi \colon D \to \mathbb{R}^k$$

with values in the $k$-dimensional vector space. We do not require that $n = k$.

Define the covariant derivative of $\psi$ by the familiar formula

$$\nabla_\alpha \psi = \frac{\partial \psi}{\partial x^\alpha} + A_\alpha \psi, \tag{10.19}$$

where $A_\alpha$ is a $k \times k$ matrix depending on the point $x \in D$, $\alpha = 1, \ldots, n$. For the definition to be independent of the choice of coordinates in $D$, we will require that the matrix-valued functions $A_\alpha$ form a tensor and the covariant derivative be written in coordinates $y = (y^1(x), \ldots, y^n(x))$ as

$$\nabla_\mu \psi = \frac{\partial \psi}{\partial y^\mu} + \tilde{A}_\mu \psi, \quad \text{where } \tilde{A}_\mu = \frac{\partial x^\alpha}{\partial y^\mu} A_\alpha.$$

However, the field $A$ may vary not only under coordinate changes in $D$, but also under basis changes in $\mathbb{R}^k$ depending on $x$ (which is the main point of our construction here).

We say that $A_\alpha$ form a *gauge field* for the group $G$ if the matrices $A_\alpha$ lie in the Lie algebra $\mathfrak{g}$ of the group $G$ and covariant differentiation commutes with basis change transformations (smooth in $x$) belonging to the group $G$:

$$\psi(x) \to g(x)\psi(x), \tag{10.20}$$

i.e., for any smooth function $g\colon D \to G$ the following identity holds:

$$\nabla_\alpha(g(x)\psi(x)) = g(x)(\nabla_\alpha\psi(x)).$$

Transformation (10.20) is called a *gauge transformation* of the field $\psi$. In the physical literature gauge fields are sometimes called *compensating fields.*

In geometry the construction just described is the general definition of a *linear connection.*

**Theorem 10.7.** *Covariant differentiation* (10.19) *commutes with transformations* (10.20) *if and only if the coefficients $A_\alpha$ change under these transformations by the formula*

$$\begin{aligned} A_\alpha(x) &= g^{-1}(x)A'_\alpha(x)g(x) + g^{-1}(x)\,\frac{\partial g(x)}{\partial x^\alpha}, \\ A'_\alpha(x) &= g(x)A_\alpha(x)g^{-1}(x) - \frac{\partial g(x)}{\partial x^\alpha}\,g^{-1}(x). \end{aligned} \tag{10.21}$$

**Proof.** We have

$$\nabla_\alpha(g\psi) = \frac{\partial(g\psi)}{\partial x^\alpha} + A'_\alpha(g\psi) \quad \text{and} \quad g(\nabla_\alpha\psi) = g\Big(\frac{\partial\psi}{\partial x^\alpha} + A_\alpha\psi\Big).$$

Equating these two expressions we obtain (10.21). □

A connection $A_\alpha$ is said to be *trivial* if there is a function $g_0\colon D \to G$ such that

$$A_\alpha(x) = g_0^{-1}(x)\,\frac{\partial g_0(x)}{\partial x^\alpha}.$$

In this case the gauge transformation $g_0(x)$, common for all fields $\psi(x)$, reduces covariant differentiation to the form

$$\nabla_\alpha\psi = \frac{\partial\psi}{\partial x^\alpha}.$$

Consider the simplest examples of gauge fields.

Example 1. Linear connection on tangent vector fields. In this case $G = \mathrm{GL}(n, \mathbb{R})$, where $n$ is the dimension of $D$. The coordinates $x^1, \dots, x^n$ specify, at each point of $D$, the basis $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ in the space

of tangent vectors. Hence tangent vector fields may be regarded as vector-functions $\xi = (\xi^1, \dots, \xi^n)$ with values in $\mathbb{R}^n$. The gauge transformations are induced by local changes of coordinates $x \to y(x)$:

$$\xi^i \to \tilde{\xi}^k = \frac{\partial y^k}{\partial x^i} \xi^i = g(x)\xi,$$

where

$$g(x) = \left(\frac{\partial y^k(x)}{\partial x^i}\right) \in \mathrm{GL}(n, \mathbb{R}).$$

The Lie algebra of the group $\mathrm{GL}(n, \mathbb{R})$ consists of all $n \times n$ matrices. Hence the connection coefficients $A_j(x)$ themselves are $n \times n$ matrices. Denote their entries by $(A_j)^k_i = \Gamma^k_{ij}$. The covariant derivative of the vector function $\xi$ (i.e., of the tangent vector field) has the form

$$\nabla_j \xi = \frac{\partial \xi}{\partial x^j} + A_j \xi, \quad (\nabla_j \xi)^k = \frac{\partial \xi^k}{\partial x^j} + \Gamma^k_{ij} \xi^i.$$

A change of coordinates generates a transformation of connection coefficients $\Gamma^k_{ij}$ by formula (10.21):

$$\Gamma^k_{ij} \to \Gamma^\nu_{\lambda j} = \frac{\partial y^\nu}{\partial x^k} \Gamma^k_{ij} \frac{\partial x^i}{\partial y^\lambda} + \frac{\partial y^\nu}{\partial x^l} \frac{\partial}{\partial x^j} \left(\frac{\partial x^l}{\partial y^\lambda}\right).$$

Taking into account that $A_j$ is a covector, i.e., $\tilde{A}_\mu = \frac{\partial x^j}{\partial y^\mu} A_j$, we obtain

$$\tilde{\Gamma}^\nu_{\lambda\mu} = \frac{\partial y^\nu}{\partial x^k} \frac{\partial x^j}{\partial y^\mu} \Gamma^k_{ij} \frac{\partial x^i}{\partial y^\lambda} + \frac{\partial y^\nu}{\partial x^l} \frac{\partial^2 x^l}{\partial y^\lambda \partial y^\mu}.$$

Thus we have derived once more formula (10.10) for transformation of Christoffel symbols under coordinate changes.

Example 2. Electromagnetic field. Let $G = \mathrm{U}(1) = \mathrm{SO}(2)$ be the one-dimensional commutative group, and let $G$ act on $\mathbb{R}^2 = \mathbb{C}$ by rotations: $e^{i\varphi}(\xi) = e^{i\varphi}\xi$, where $\xi \in \mathbb{C}$ and $e^{i\varphi} \in \mathrm{U}(1)$, $\varphi \in \mathbb{R}$. The Lie algebra $\mathfrak{g}$ consists of purely imaginary numbers, so the connection has the form

$$\nabla_\alpha = \frac{\partial}{\partial x^\alpha} + ieA_\alpha,$$

where the $A_\alpha$ are real-valued (they form the potential of the electromagnetic field) and $e$ is a physical quantity (charge). This is how the vector potential of electromagnetic field enters into the quantum mechanics equations — similarly to covariant derivatives in geometry, as observed by H. Weyl. The gauge transformation $\psi \to e^{i\varphi(x)}\psi$ generates the following transformation of the connection:

$$A_\alpha \to A_\alpha + \frac{\partial \varphi}{\partial x^\alpha}.$$

What makes the gauge fields so popular, is their important role in physics of elementary particles.

Let a vector-function $\psi$ describe some physical field with interior degrees of freedom (their number equals the number of components of the vector-function). Suppose that such fields describe physical states not in one-to-one way, but up to a gauge transformation, i.e., vector-functions $\psi(x)$ and $g(x)\psi(x)$ describe the same physical state. Then, when trying to build a mathematical theory of such fields, one must define differentiation so as to preserve this invariance. This gives rise to connections $A_\alpha$, which themselves describe fields interacting with the "material fields" $\psi$ (see 15.3.4) and provide a natural extension of the electromagnetic field. These fields describe some fundamental particles, and they are called the *Yang–Mills fields*.

**10.1.4. Cartan connections.** Recall that the group of all affine transformations of the Cartesian space $\mathbb{R}^k$ is denoted by $\mathrm{A}(k,\mathbb{R})$. An affine transformation of $\mathbb{R}^k$ is uniquely specified as $\xi \to A\xi + b$, where $A \in \mathrm{GL}(k,\mathbb{R})$, $b \in \mathbb{R}^k$, and any transformation of this form is affine. This group is also a matrix group, being imbedded into $\mathrm{GL}(k+1,\mathbb{R})$:

$$C = (A, b) \to \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}. \tag{10.22}$$

We fix a hyperplane $\xi^{k+1} = 1$ in $\mathbb{R}^{k+1}$ and assign to each vector $\xi \in \mathbb{R}^k$ the point on this hyperplane by the following rule: $\xi = (\xi^1, \dots, \xi^k) \to (\xi, 1) = (\xi^1, \dots, \xi^k, 1)$. Then

$$\begin{pmatrix} A\xi + b \\ 1 \end{pmatrix} = \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ 1 \end{pmatrix}.$$

Let $D$ be a domain in $\mathbb{R}^n$ with coordinates $x^1, \dots, x^n$. Let $\hat{A}_\alpha(x)$ be an array of functions with values in the Lie algebra of the group $\mathrm{A}(k,\mathbb{R})$ specifying in the domain $D$ covariant differentiation of $\mathbb{R}^k$-valued vector-functions $\psi$:

$$\hat{\nabla}_\alpha \psi = \frac{\partial \psi}{\partial x^\alpha} + \hat{A}_\alpha \psi.$$

We will require that the functions $\hat{A}_\alpha$ form a covector relative to changes of coordinates in $D$ (i.e., that $\hat{A}_\alpha\, dx^\alpha$ specify a matrix-valued 1-form on $D$) and that covariant differentiation be invariant with respect to gauge transformations

$$\psi(x) \to g(x)\psi(x)$$

with values in a group $G \subset \mathrm{A}(k,\mathbb{R})$. This means that

$$g(x)\big(\hat{\nabla}_\alpha \psi(x)\big) = \hat{\nabla}_\alpha (g(x)\psi(x)), \quad g(x) \in G \subset \mathrm{A}(k,\mathbb{R}).$$

Such an array of functions $\hat{A}_\alpha(x)$ is called an *affine connection.*

Here we make an important remark: since multiplication of a vector by a number or summation of two vectors are not invariant with respect to affine transformations, the covariant differentiation $\hat{\nabla}_\alpha$ is not linear in $\psi$.

There is a simple construction, introduced by Cartan, that associates with each linear connection $A_\alpha$ an affine connection on the same space of vector-functions. This construction uses the imbedding (10.22). Define the affine connection as the restriction of the linear connection on $\mathbb{R}^{k+1}$-valued functions to vector-functions of the form $(\psi, 1)$:

$$\hat{\nabla}_\alpha(\psi, 1) = \Big(\frac{\partial \psi}{\partial x^\alpha} + A_\alpha \psi + \varphi_\alpha, 1\Big),$$

where $\varphi_\alpha(x) \in \mathbb{R}^k$. Set the vectors $\varphi_\alpha$ to be

$$\varphi^\mu_\alpha(x) = \delta^\mu_\alpha, \quad \alpha, \mu = 1, \dots, k.$$

Thus we obtain the affine connection

$$(\hat{\nabla}_\alpha \psi)^\mu = (\nabla_\alpha \psi)^\mu + \delta^\mu_\alpha, \tag{10.23}$$

where $\nabla_\alpha$ is the given linear connection. The affine connection (10.23) is called the *Cartan connection*.

The gauge transformations $\psi(x) \to g(x)\psi(x) + b(x)$, where $g \in \mathrm{GL}(k, \mathbb{R})$, $b \in \mathbb{R}^k$, do not preserve the equality $\varphi^\mu_\alpha = \delta^\mu_\alpha$: the connection transforms by the rules

$$A_\alpha \to gA_\alpha g^{-1} - \frac{\partial g}{\partial x^\alpha} g^{-1}, \quad \varphi_\alpha \to g\varphi_\alpha - \frac{\partial b}{\partial x^\alpha} - \Big(gA_\alpha g^{-1} - \frac{\partial g}{\partial x^\alpha} g^{-1}\Big) b.$$

**10.1.5. Parallel translation.** The notion of parallel lines in Euclidean geometry is based on the fifth of Euclid's postulates:

*Given a straight line in the Euclidean plane, for any point $P$ there is a unique line passing through $P$ and parallel to the given line.*

Here the parallel line either coincides with the given line (when $P$ lies on it) or does not intersect it. In Euclidean coordinates this manifests itself in that one can choose tangent vectors to these lines so that they coincide as vectors of two-dimensional space; these vectors (originating from different points) are called parallel. This approach allows us to extend the notion of parallelism to multidimensional Euclidean spaces.

In general we cannot introduce global coordinates on the entire space (e.g., on a sphere) and reduce parallelism to arithmetic equality between coordinates of vectors. In order to define parallelism we will employ another property of parallel vectors in Euclidean spaces.

Let $P$ and $Q$ be two points on the plane, and let $x(t)$ be a smooth curve going from $x(0) = P$ to $x(1) = Q$. Consider the equation for tangent vectors to the curve $x(t)$:

$$\frac{dv}{dt} = 0 \tag{10.24}$$

with initial condition $v(0) = v_0$. This is a first-order ordinary differential equation which has a unique solution. In Euclidean coordinates equation (10.24) takes the simple form

$$\frac{dv^i}{dt} = 0;$$

hence the vectors $v(t)$ tangent to the curve at the points $x(t)$ will be parallel to $v_0$. In particular, the vector $v(1)$ at the point $Q$ is parallel to the vector $v_0$ at the point $P$. In curvilinear coordinates equation (10.24) rewrites as

$$\nabla_{\dot{x}} v = 0,$$

and, since covariant differentiation takes tensors into tensors, it has the same solutions as (10.24).

This approach underlies the definition of parallel translation in arbitrary spaces with connection $\nabla$. Namely, we will introduce connections also on manifolds different from domains in Euclidean space. To this end we cover a manifold by an atlas, and in each of its charts introduce a connection as in a domain of the Euclidean space. We only have to require that the connections on intersections of the charts be compatible. This means that the symbols $\Gamma^i_{jk}$ taken relative to different coordinates $(x)$ and $(z)$ must be related by formulas (10.10). Then the covariant derivatives of tensors are well defined on the entire manifold.

A tensor field $T$ is said to be *parallel* along a curve $x(t)$, $a \leq t \leq b$, if it satisfies the equation

$$\nabla_{\dot{x}} T = \dot{x}^i \nabla_i T = 0. \tag{10.25}$$

The tensor $T(b)$ at the point $x(b)$ is called the *parallel translation* of the tensor $T(a)$ from the point $x(a)$ along the curve $x(t)$.

In particular, for vector fields equation (10.25) takes a simple form

$$\nabla_{\dot{x}} T^i = \frac{dT^i}{dt} + \dot{x}^k \Gamma^i_{jk} T^j = 0, \quad i = 1, \ldots, n.$$

Equation (10.25) is of first order and linear in $T$. By the theorem on the existence and uniqueness of solutions of differential equations we obtain the following fact.

**Theorem 10.8.** *Let $x(t)$, $a \leq t \leq b$, be a smooth curve in a space with connection $\nabla$. Then the parallel translation of a tensor $T$ along the curve $x(t)$ is well defined, and its outcome is uniquely defined by the initial value $T(a)$ and depends on it linearly.*

In Euclidean geometry the outcome of parallel translation of a tensor from a point $P$ to a point $Q$ does not depend on the curve joining these

points. This property singles out this geometry, since in general, parallel translation depends on the curve.

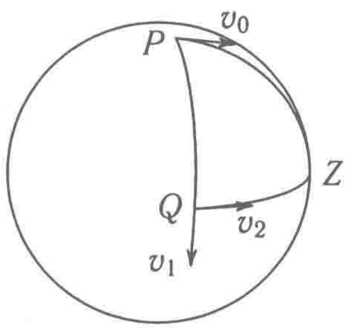


**Figure 10.1.** Parallel translation on the sphere.

EXAMPLE. Let $S^2$ be the unit sphere in Euclidean space with coordinates $x^1, x^2, x^3$. Define covariant differentiation on it by formula (10.8) (as on an imbedded surface). Take the points $P = (0,0,1)$, $Q = (1,0,0)$, and $Z = (0,1,0)$ on the sphere. Join these points by the arcs of great circles:

$$x_1(t) = \left(0, \sin\frac{\pi t}{2}, \cos\frac{\pi t}{2}\right), \quad x_2(t) = \left(\sin\frac{\pi t}{2}, \cos\frac{\pi t}{2}, 0\right),$$
$$x_3(t) = \left(\sin\frac{\pi t}{2}, 0, \cos\frac{\pi t}{2}\right), \quad 0 \le t \le 1.$$

Then 1) $x_1(0) = P$, $x_1(1) = Z$; 2) $x_2(0) = Z$, $x_2(1) = Q$; 3) $x_3(0) = P$, $x_3(1) = Q$. At the point $P$ take the tangent vector $v_0 = (0,1,0)$. By a parallel translation we carry it along $x_1(t)$ into the point $Z$ to obtain the vector $w$, next we carry the vector $w$ along $x_2(t)$ to obtain the vector $v_1$ at the point $Q$. Then we define the vector $v_2$ at the point $Q$ by the parallel translation of $v_0$ along $x_3(t)$. As a result, we obtain

$$v_1 = w = (0,0,-1), \quad v_2 = (0,1,0),$$

and we see that $v_1 \neq v_2$ (Figure 10.1).

The difference between parallel translations along different curves is due to curvature of the space, which will be considered in Section 10.2.

**10.1.6. Connections compatible with a metric.** The concepts of a connection and a Riemannian metric are unrelated to each other. At the same time we actually use two different definitions of Euclidean coordinates:

1) coordinates in which the metric has the form $g_{ij} = \delta_{ij}$, and

2) coordinates in which the connection is trivial: $\Gamma^i_{jk} \equiv 0$.

In fact, these are different notions, but there is a canonical correspondence between metrics and connections such that the Euclidean metric

$g_{ij} = \delta_{ij}$ corresponds to the Euclidean connection $\Gamma^i_{jk} \equiv 0$. Now we will explain how this correspondence is established.

First of all we note that for Euclidean coordinates the operator of parallel translation is an isometry. This means the following.

Let $x(t)$, $a \le t \le b$, be a curve in the space. To each $t$ corresponds the linear operator $L_t$ that associates with the tangent vector at $x(a)$ the result of its parallel translation to the point $x(t)$ along this curve. In Euclidean space this operator preserves the scalar product:

$$\langle \xi, \eta \rangle = \langle L_t \xi, L_t \eta \rangle. \tag{10.26}$$

In the general case, if the parallel translation of vectors is an isometry for any curve, i.e., (10.26) holds, the connection is said to be *compatible with the metric.*

Now we will find analytic conditions describing such connections. Let $\xi(t)$, $\eta(t)$ be a pair of parallel vector fields along a curve $x(t)$. Suppose that the connection is compatible with the metric:

$$\frac{d}{dt}\langle \xi, \eta \rangle = \frac{d}{dt}(g_{ij}(x(t))\, \xi^i(t)\, \eta^j(t)) = 0.$$

We rewrite this relation in terms of covariant derivatives:

$$\frac{d}{dt}(g_{ij}(x(t))\, \xi^i(t)\, \eta^j(t)) = (\nabla_{\dot{x}} g_{ij})\xi^i \eta^j + g_{ij}(\nabla_{\dot{x}} \xi^i)\eta^j + g_{ij}\xi^i(\nabla_{\dot{x}} \eta^j) = 0.$$

Since the fields $\xi$ and $\eta$ are parallel, we conclude that at each point

$$(\nabla_{\dot{x}} g_{ij})\xi^i \eta^j = 0. \tag{10.27}$$

Note that any vector $\dot{x}$ is a velocity vector of some curve, and any vector $\xi$ or $\eta$ can be extended to a parallel vector field along this curve. Hence (10.27) implies that

$$\nabla_k g_{ij} \equiv 0, \quad i, j, k = 1, \ldots, n. \tag{10.28}$$

Thus we arrive at an equivalent definition: a connection $\Gamma^k_{ij}$ is said to be *compatible with the metric* $g_{ij}$ if the covariant derivative of the metric tensor is identically equal to zero, i.e., equalities (10.28) hold.

We do not require the metric to be Riemannian, since this definition is suitable also for pseudo-Riemannian and degenerate metrics.

**Lemma 10.3.** *Let $\nabla$ be a connection compatible with a metric $g_{ij}$. Then:*

1. *For $\xi(t)$ and $\eta(t)$ vector fields along the curve $x(t)$,*

$$\frac{d}{dt}\langle \xi, \eta \rangle = \langle \nabla_{\dot{x}} \xi, \eta \rangle + \langle \xi, \nabla_{\dot{x}} \eta \rangle.$$

2. *Lowering of any tensor superscript commutes with covariant differentiation.*

**Proof.** 1. It follows from the Leibniz formula for covariant differentiation that

$$\frac{d}{dt}\langle\xi,\eta\rangle = (\nabla_{\dot{x}}g_{ij})\xi^i\eta^j + g_{ij}(\nabla_{\dot{x}}\xi^i)\eta^j + g_{ij}\xi^i(\nabla_{\dot{x}}\eta^j) = \langle\nabla_{\dot{x}}\xi,\eta\rangle + \langle\xi,\nabla_{\dot{x}}\eta\rangle.$$

2. In a similar way, we have

$$\nabla_k(g_{lm}T^{m(i)}_{(j)}) = (\nabla_k g_{lm})T^{m(i)}_{(j)}) + g_{lm}(\nabla_k T^{m(i)}_{(j)}) = g_{lm}(\nabla_k T^{m(i)}_{(j)}).$$

Hence the lemma. □

The Euclidean connection (10.5) and the connection on an imbedded surface (10.7), which started our discussion, have an additional property of symmetry. Namely, a connection $\Gamma^i_{jk}$ is said to be *symmetric* if

$$\Gamma^i_{jk} = \Gamma^i_{kj}, \quad i,j,k = 1,\dots,n,$$

or, equivalently, if the torsion tensor vanishes, $T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj} = 0$.

Now we state an important theorem.

**Theorem 10.9.** *Let $g_{ij}$ be a nondegenerate metric (i.e., $g = \det(g_{ij}) \neq 0$). Then there exists a unique symmetric connection compatible with this metric $g_{ij}$. In any coordinate system $(x^1,\dots,x^n)$ this connection is specified by the Christoffel formulas*

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{lj}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right). \tag{10.29}$$

**Proof.** Since the connection is compatible with the metric, we have

$$\nabla_k g_{ij} = \frac{\partial g_{ij}}{\partial x^k} - \Gamma^l_{ik}g_{lj} - \Gamma^l_{jk}g_{il} = 0. \tag{10.30}$$

Let us solve this system of equations for $\Gamma^k_{ij}$. On lowering the superscript,

$$\Gamma_{kij} = g_{kl}\Gamma^l_{ij},$$

equations (10.30) become

$$\Gamma_{ijk} + \Gamma_{jik} = \frac{\partial g_{ij}}{\partial x^k}.$$

Since the connection is symmetric, $\Gamma_{ijk} = \Gamma_{ikj}$ for all $i,j,k$. By cyclic permutations of indices $i,j,k$ we obtain the equations

$$\Gamma_{ijk} + \Gamma_{jik} = \frac{\partial g_{ij}}{\partial x^k}, \quad \Gamma_{jki} + \Gamma_{kji} = \frac{\partial g_{kj}}{\partial x^i}, \quad \Gamma_{ikj} + \Gamma_{kij} = \frac{\partial g_{ik}}{\partial x^j}.$$

Let (1), (2), (3) be the left-hand sides of these formulas. Then symmetry of the connection implies that (2) + (3) − (1) = $2\Gamma_{kij}$. Therefore

$$\Gamma_{kij} = \frac{1}{2}\left(\frac{\partial g_{kj}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right) = g_{kl}\Gamma^l_{ij}.$$

Now we raise the subscript $k$ to obtain (10.29). The proof is completed. □

**Corollary 10.3.** *If in some coordinates all the first order derivatives of $g_{ij}$ vanish at some point, then the Christoffel symbols of the symmetric connection $\Gamma^k_{ij}$ compatible with the metric vanish at this point.*

The covariant divergence of a vector field $v$ is defined as

$$\operatorname{div} v^i = \nabla_i v^i = v^i_{;i}.$$

**Corollary 10.4.** *If the connection is symmetric and compatible with a non-degenerate metric $g_{ij}$, then the covariant divergence of the vector field is*

$$\nabla_i v^i = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^i}(\sqrt{|g|} v^i), \tag{10.31}$$

*where* $g = \det(g_{ij})$.

In order to obtain (10.31), substitute the expression

$$\Gamma^i_{ki} = \frac{1}{2} g^{il}\Big(\frac{\partial g_{lk}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^k} - \frac{\partial g_{ki}}{\partial x^l}\Big) = \frac{1}{2} g^{il} \frac{\partial g_{il}}{\partial x^k} = \frac{1}{2g}\frac{\partial g}{\partial x^k} = \frac{\partial}{\partial x^k}\log(\sqrt{|g|})$$

into the equality $\nabla_i v^i = \frac{\partial v^i}{\partial x^i} + \Gamma^i_{ki} v^k$. We obtain

$$\nabla_i v^i = \frac{\partial v^i}{\partial x^i} + \frac{1}{2g}\frac{\partial g}{\partial x^k} v^k = \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^i}(\sqrt{|g|}v^i).$$

In Euclidean space the meaning of divergence is as follows. Let

$$x^i \to x^i + v^i t$$

be a small displacement of points of the space. Then the volume element of a domain transforms under this displacement by the formula

$$dx^1 \wedge \cdots \wedge dx^n \to (1 + v^i_{,i} t + o(t))\, dx^1 \wedge \cdots \wedge dx^n.$$

In a space with Riemannian metric the volume element is equal to $\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n$, and under the displacement it becomes

$$\Big(\sqrt{g} + \Big(\frac{1}{2\sqrt{g}}\frac{\partial g}{\partial x^k} v^k + v^i_{,i}\sqrt{g}\Big)t + o(t)\Big)\, dx^1 \wedge \cdots \wedge dx^n,$$

i.e., as in the Euclidean case, the increment of the volume, up to quantities of order $o(t)$, equals

$$\operatorname{div} v(\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n)t.$$

We conclude that the divergence has its usual form $\operatorname{div} v = v^i_{,i}$ only if the volume element in the given coordinates coincides with the Euclidean element, i.e., $g \equiv 1$.

## 10.2. Curvature tensor

**10.2.1. Definition of the curvature tensor.** In 10.1.5 we have already given an example where the result of parallel translation of a vector on the unit sphere depends on the path rather than only on the endpoints. In particular, parallel translation of a vector along a closed path may result in a vector different from the initial vector.

This is not the case in Euclidean geometry: parallel translation is specified by the equation

$$\frac{d\xi}{dt} = 0,$$

and its result depends only on the endpoints. This is due to the fact that in Euclidean geometry the covariant derivatives commute:

$$\nabla_j \xi^i = \frac{\partial X^i}{\partial x^j}, \quad (\nabla_j \nabla_k - \nabla_k \nabla_j)\xi = 0.$$

However, this property fails for a general connection. Recall that the parallel translation of a vector field $\xi$ along a path $x(t)$ is determined by the equation

$$\frac{d\xi^i}{dt} + \Gamma^i_{kj}\xi^k \dot{\xi}^j = 0.$$

We have

$$\begin{aligned}
\nabla_j \nabla_k \xi^i &= \nabla_j \Big(\frac{\partial \xi^i}{\partial x^k} + \Gamma^i_{lk}\xi^l\Big) \\
&= \frac{\partial}{\partial x^j}\Big(\frac{\partial \xi^i}{\partial x^k} + \Gamma^i_{lk}\xi^l\Big) + \Gamma^i_{mj}\Big(\frac{\partial \xi^m}{\partial x^k} + \Gamma^m_{lk}\xi^l\Big) - \Gamma^m_{kj}\Big(\frac{\partial \xi^i}{\partial x^m} + \Gamma^i_{lm}\xi^l\Big) \\
&= \frac{\partial^2 \xi^i}{\partial x^j \partial x^k} + \frac{\partial \xi^l}{\partial x^j}\,\Gamma^i_{lk} + \Gamma^i_{mj}\,\frac{\partial \xi^m}{\partial x^k} - \Gamma^m_{kj}\,\frac{\partial \xi^i}{\partial x^m} \\
&\quad + \xi^l\,\frac{\partial \Gamma^i_{lk}}{\partial x^j} + \Gamma^i_{mj}\,\Gamma^m_{lk}\,\xi^l - \Gamma^m_{kj}\,\Gamma^i_{lm}\xi^l.
\end{aligned}$$

This implies that the commutator of covariant derivatives in the directions of basis vector fields equals

$$\begin{aligned}
(\nabla_k \nabla_j - \nabla_j \nabla_k)\xi^i = \Big(\frac{\partial \Gamma^i_{lj}}{\partial x^k} - \frac{\partial \Gamma^i_{lk}}{\partial x^j}\Big)\xi^l + (\Gamma^i_{mk}\Gamma^m_{lj} - \Gamma^i_{mj}\Gamma^m_{lk})\xi^l \\
- (\Gamma^m_{jk} - \Gamma^m_{kj})\frac{\partial \xi^i}{\partial x^m} - (\Gamma^m_{jk} - \Gamma^m_{kj})\Gamma^i_{lm}\xi^l.
\end{aligned}$$

Thus we have proved the following theorem.

**Theorem 10.10.** *In arbitrary coordinates $x^1, \dots, x^n$ the commutator of covariant differentiations equals*

$$(\nabla_k \nabla_j - \nabla_j \nabla_k)\xi^i = R^i_{lkj}\xi^l + T^m_{kj}\nabla_m \xi^i, \tag{10.32}$$

*where*

$$R^i_{lkj} = \frac{\partial \Gamma^i_{lj}}{\partial x^k} - \frac{\partial \Gamma^i_{lk}}{\partial x^j} + \Gamma^i_{mk}\Gamma^m_{lj} - \Gamma^i_{mj}\Gamma^m_{lk},$$
$$T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj}.$$

As will be shown below, the quantities $R^i_{ljk}$ constitute a tensor, which is called the *curvature tensor*, or the *Riemann tensor*. We have already shown in 10.1.1 that $T^i_{jk}$ is a tensor, called the torsion tensor (Lemma 10.1).

Recall that a connection is said to be symmetric if its torsion tensor vanishes,

$$T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj} = 0.$$

Since

$$\nabla_j e_k = \Gamma^i_{kj} e_i,$$

where $e_1, \ldots, e_n$ are basis vector fields, symmetry of a connection is equivalent to

$$\nabla_j e_k = \nabla_k e_j \qquad \text{for} \quad j, k = 1, \ldots, n.$$

Thus we obtain the following result.

**Corollary 10.5.** *If a connection is symmetric, then*

$$(\nabla_k \nabla_j - \nabla_j \nabla_k)\xi^i = R^i_{lkj}\xi^l.$$

Recall that by Lemma 10.2, for a symmetric connection, in a neighborhood of any given point, one can introduce coordinates in which all the Christoffel symbols at this point vanish. In this case the Riemann tensor at this point has also a simple form:

$$R^i_{lkj} = \frac{\partial \Gamma^i_{lj}}{\partial x^k} - \frac{\partial \Gamma^i_{lk}}{\partial x^j}. \tag{10.33}$$

The following lemma implies that $R^i_{ljk}$ is a tensor.

**Lemma 10.4.** *For arbitrary vector fields $\xi$, $\eta$, $\zeta$,*

$$(\nabla_\xi \eta - \nabla_\eta \xi - [\xi, \eta])^i = T^i_{jk}\eta^j\xi^k = -T^i_{kj}\xi^k\eta^j, \tag{10.34}$$

$$(\nabla_\eta \nabla_\xi \zeta - \nabla_\xi \nabla_\eta \zeta + \nabla_{[\xi,\eta]}\zeta)^i = R^i_{lkj}\xi^j\eta^k\zeta^l, \tag{10.35}$$

*where $[\xi, \eta]$ is the commutator of vector fields and the left-hand sides are the* $i$th *components of the fields in parentheses. In particular, the values of the left-hand sides of these formulas at a point are determined only by the values of the vector fields $\xi$, $\eta$, and $\zeta$ at this point.*

For basis vector fields $\xi = e_j$ and $\eta = e_k$, their commutator vanishes and formula (10.35) coincides with (10.32).

**Proof.** Denote, for brevity,

$$T(\xi,\eta) = \nabla_\xi\eta - \nabla_\eta\xi - [\xi,\eta], \quad R(\xi,\eta)\zeta = \nabla_\eta\nabla_\xi\zeta - \nabla_\xi\nabla_\eta\zeta + \nabla_{[\xi,\eta]}\zeta.$$

These expressions are linear with respect to the fields $\xi$, $\eta$, and $\zeta$. Therefore, in order to reduce the proof of the identities (10.34) and (10.35) to their verification for basis vector fields, we must prove that $T(\xi,\eta)$ and $R(\xi,\eta)\zeta$ are linear not only under multiplication of the fields by constants, but also under their multiplication by smooth functions.

Let $f$ be a smooth function. We will denote by $\partial_\xi f$ the directional derivative of $f$ in the direction of the vector field $\xi$.

We have

$$\begin{aligned} T(f\xi,\eta) &= \nabla_{f\xi}\eta - \nabla_\eta(f\xi) - [f\xi,\eta] \\ &= f[\nabla_\xi\eta - \nabla_\eta\xi - [\xi,\eta]] - (\partial_\eta f)\xi + (\partial_\eta f)\xi = fT(\xi,\eta). \end{aligned}$$

In a similar manner one shows that $T(\xi,f\eta) = fT(\xi,\eta)$.

For $R(\xi,\eta)\zeta$ we show by a direct calculation that

$$\begin{aligned} R(f\xi,\eta)\zeta &= (\nabla_\eta\nabla_{f\xi} - \nabla_{f\xi}\nabla_\eta + \nabla_{[f\xi,\eta]})\zeta \\ &= (f\nabla_\eta\nabla_\xi + (\partial_\eta f)\nabla_\xi - f\nabla_\xi\nabla_\eta + \nabla_{f[\xi,\eta]-(\partial_\eta f)\xi})\zeta = fR(\xi,\eta)\zeta, \end{aligned}$$

and, since $R(\xi,\eta)\zeta = -R(\eta,\xi)\zeta$, we have $R(\xi,f\eta) = fR(\xi,\eta)$.

Let us show that $R(\xi,\eta)(f\zeta) = fR(\xi,\eta)\zeta$. This is also done by a direct calculation:

$$\begin{aligned} R(\xi,\eta)(f\zeta) &= \nabla_\eta\nabla_\xi(f\zeta) - \nabla_\xi\nabla_\eta(f\zeta) + \nabla_{[\xi,\eta]}(f\zeta) \\ &= \nabla_\eta[(\partial_\xi f)\zeta + f\nabla_\xi\zeta] - \nabla_\xi[(\partial_\eta f)\zeta + f\nabla_\eta\zeta] \\ &\quad + (\partial_{[\xi,\eta]}f)\zeta + f\nabla_{[\xi,\eta]}\zeta \\ &= (\partial_\eta\partial_\xi f)\zeta + (\partial_\xi f)\nabla_\eta\zeta + (\partial_\eta f)\nabla_\xi\zeta + f\nabla_\eta\nabla_\xi\zeta \\ &\quad - (\partial_\xi\partial_\eta f)\zeta - (\partial_\eta f)\nabla_\xi\zeta - (\partial_\xi f)\nabla_\eta\zeta - f\nabla_\xi\nabla_\eta\zeta \\ &\quad + (\partial_\xi\partial_\eta f - \partial_\eta\partial_\xi f)\zeta + f\nabla_{[\xi,\eta]}\zeta \\ &= fR(\xi,\eta)\zeta. \end{aligned}$$

Now it remains to prove equalities (10.34) and (10.35) for basis vector fields $\xi = e_k$, $\eta = e_l$, and $\zeta = e_j$. By the definition of the Christoffel symbols,

$$\nabla_{e_j} e_k = \Gamma^i_{kj} e_i.$$

This implies the relation

$$\nabla_{e_j} e_k - \nabla_{e_k} e_j - [e_j, e_k] = (\Gamma^i_{kj} - \Gamma^i_{jk}) e_i,$$

which is equivalent to the equality

$$T(e_j, e_k) = \nabla_{e_j} e_k - \nabla_{e_k} e_j.$$

Now we write

$$\nabla_{e_k}\nabla_{e_j}e_l - \nabla_{e_j}\nabla_{e_k}e_l + \nabla_{[e_j,e_k]}e_l = \nabla_{e_k}(\Gamma^i_{lj}e_i) - \nabla_{e_j}(\Gamma^i_{lk}e_i)$$
$$= \frac{\partial\Gamma^i_{lj}}{\partial x^k} + \Gamma^m_{lj}\Gamma^i_{mk} - \frac{\partial\Gamma^i_{lk}}{\partial x^j} - \Gamma^m_{lk}\Gamma^i_{mj} = R^i_{lkj}e_i = R(e_j,e_k)e_l.$$

This proves the lemma. □

**Corollary 10.6.** *A necessary condition for a connection to be reducible to the form*

$$\Gamma^i_{jk} = 0 \qquad \textit{for} \quad i,j,k = 1,\dots,n,$$

*in the entire domain of an $n$-dimensional manifold is that the Riemann tensor in this domain vanishes:*

$$R^i_{lkj} = 0 \qquad \textit{for} \quad i,j,k = 1,\dots,n.$$

**Corollary 10.7** (Bianchi identity). *For a symmetric connection the following equality holds:*

$$\nabla_l R^i_{mkj} + \nabla_j R^i_{mlk} + \nabla_k R^i_{mjl} = 0 \tag{10.36}$$

*for all $i,j,k,l = 1,\dots,n$.*

**Proof.** In a neighborhood of an arbitrary point $x_0 \in M^n$, we introduce coordinates in which all the Christoffel symbols at this point vanish (Lemma 10.2). In these coordinates

$$R^i_{lkj} = \frac{\partial\Gamma^i_{lj}}{\partial x^k} - \frac{\partial\Gamma^i_{lk}}{\partial x^j},$$

and covariant differentiation at this point coincides with the ordinary differentiation. Now we rewrite the left-hand side of (10.36) as

$$\Big(\frac{\partial^2\Gamma^i_{mj}}{\partial x^k\partial x^l} - \frac{\partial^2\Gamma^i_{mk}}{\partial x^j\partial x^l}\Big) + \Big(\frac{\partial^2\Gamma^i_{mk}}{\partial x^l\partial x^j} - \frac{\partial^2\Gamma^i_{ml}}{\partial x^k\partial x^j}\Big) + \Big(\frac{\partial^2\Gamma^i_{ml}}{\partial x^j\partial x^k} - \frac{\partial^2\Gamma^i_{mj}}{\partial x^l\partial x^k}\Big) = 0.$$

Hence the corollary. □

The Bianchi identity for arbitrary connections will be treated in 15.3.3.

**10.2.2. Symmetries of the curvature tensor.** It follows from the very definition of the curvature tensor that it is skew-symmetric in the last two indices. When the connection is symmetric or compatible with metric, the curvature tensor possesses additional symmetries to be listed in the following theorem.

Before doing that, for the case of a manifold equipped with a metric $g_{ij}$, we define the tensor $R_{iljk}$ obtained by lowering the superscript from the curvature tensor:

$$R_{mlkj} = g_{mi}R^i_{lkj}.$$

Obviously, we have

$$\langle R(X,Y)Z, W\rangle = R_{mlkj}W^m Z^l Y^k X^j.$$

**Theorem 10.11.** *Let $R^i_{lkj}$ be the curvature tensor of an affine connection on a manifold. Then:*

1) *The following equality holds:*

$$R^i_{ljk} = -R^i_{lkj}, \quad \text{or} \quad R(X,Y)Z = -R(Y,X)Z. \tag{10.37}$$

2) *If the connection is symmetric, then*

$$R^i_{ljk} + R^i_{jkl} + R^i_{klj} = 0, \tag{10.38}$$

*or*

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0.$$

3) *If the connection is compatible with the metric $g_{ij}$, then*

$$R_{iljk} = -R_{lijk}, \quad \text{or} \quad \langle R(X,Y)Z, W\rangle = -\langle R(X,Y)W, Z\rangle. \tag{10.39}$$

4) *If the connection is symmetric and compatible with the metric $g_{ij}$ (not necessarily positive definite), then*

$$R_{iljk} = R_{jkil}, \quad \text{or} \quad \langle R(X,Y)Z, W)\rangle = \langle R(Z,W)X, Y\rangle. \tag{10.40}$$

**Proof.** Statement 1) is obvious.

To prove statement 2), it suffices to show that

$$[\nabla_j, \nabla_k]e_l + [\nabla_k, \nabla_l]e_j + [\nabla_l, \nabla_j]e_k = 0.$$

Since the connection is symmetric, $\nabla_i e_j = \nabla_j e_i$ for any pair of basis vector fields $e_i$ and $e_j$, and the left-hand side becomes

$$(\nabla_j\nabla_k e_l - \nabla_k\nabla_l e_j) + (\nabla_k\nabla_l e_j - \nabla_l\nabla_j e_k) + (\nabla_l\nabla_j e_k - \nabla_j\nabla_k e_l) = 0.$$

Statement 3) is equivalent to

$$\langle R(X,Y)Z, Z\rangle = 0$$

for any vector fields $X$, $Y$, and $Z$. Since $R^i_{jkl}$ is a tensor, for the proof of this fact we may restrict ourselves to the case where $X$ and $Y$ are basis fields. We have

$$\langle R(e_j, e_k)Z, Z\rangle = \langle[\nabla_k, \nabla_j]Z, Z\rangle = -\langle[\nabla_j, \nabla_k]Z, Z\rangle.$$

Since the connection is compatible with the metric, on differentiating the function $\langle Z, Z\rangle$ we obtain

$$\frac{1}{2}\frac{\partial^2}{\partial x^j \partial x^k}\langle Z, Z\rangle = \langle\nabla_j\nabla_k Z, Z\rangle + \langle\nabla_k Z, \nabla_j Z\rangle,$$
$$\frac{1}{2}\frac{\partial^2}{\partial x^k \partial x^j}\langle Z, Z\rangle = \langle\nabla_k\nabla_j Z, Z\rangle + \langle\nabla_j Z, \nabla_k Z\rangle.$$

Subtract the second equality from the first to obtain

$$\langle [\nabla_j, \nabla_k] Z, Z\rangle = 0,$$

which was to be proved.

Now we prove statement 4). In terms of the tensor $R_{ijkl}$ relation (10.37) becomes

$$R_{ijkl} = -R_{ijlk},$$

and (10.38) can be rewritten as

$$R_{iljk} + R_{ijkl} + R_{iklj} = R_{i[jkl]} = 0,$$

where the notation $[jkl]$ means summation over all cyclic permutations of these indices. Taking into account (10.37), (10.38), and (10.39), we obtain

$$R_{i[jlk]} + R_{l[ijk]} - R_{j[ikl]} - R_{k[ilj]} = 2(R_{lijk} - R_{jkli}) = 0,$$

which implies statement 4). The proof is completed. □

**10.2.3. The Riemann tensors in small dimensions, the Ricci tensor, scalar and sectional curvatures.** We consider a connection which is symmetric and compatible with the metric. Recall that the metric need not be positive definite.

Theorem 10.11 implies that the Riemann tensor $R_{ijkl}$ is skew-symmetric in the first and second pairs of indices,

$$R_{ijkl} = -R_{jikl} = -R_{ijlk},$$

symmetric relative to permutations of these pairs of indices,

$$R_{ijkl} = R_{klij},$$

and satisfies the relation

$$R_{i[jkl]} = R_{ijkl} + R_{iklj} + R_{iljk} = 0.$$

Since the Riemann tensor satisfies quite a number of relations, in small dimensions it is completely determined by the parameters involved in the *Ricci tensor* $R_{kl}$, which is the contraction of the curvature tensor:

$$R_{kl} = R^i_{kil}.$$

In coordinates, the Ricci tensor is given by the formula

$$R_{kl} = \frac{\partial \Gamma^i_{kl}}{\partial x^i} - \frac{\partial \Gamma^i_{ki}}{\partial x^l} + \Gamma^i_{kl}\Gamma^m_{im} - \Gamma^i_{km}\Gamma^m_{il}.$$

Relations (10.37) to (10.40) imply that the Ricci tensor is symmetric,

$$R_{ik} = R_{ki}.$$

It may also be defined as the trace of the linear mapping $\Phi_{X,Y}$:

$$\Phi_{X,Y}(\xi) = R(X,Y)\xi, \quad R_{ik}X^iY^k = \mathrm{Tr}(\Phi_{X,Y}).$$

The trace of the Ricci tensor

$$R = R^i_i = g^{ij} R_{ij}$$

is called the *scalar curvature.*

When speaking about the Ricci curvature or the scalar curvature of a manifold with a metric, we presume that the unique symmetric connection compatible with the metric has been chosen on the manifold.

Now we examine how many independent components the Riemann tensor may have in case of manifolds of a small dimension.

1. TWO-DIMENSIONAL CASE: $\dim M = 2$. It follows from the symmetries of the Riemann tensor that the entire tensor is determined by a single component $R_{1212} = -R_{2112}$. All the other components either vanish, or can be obtained from this one by permutations. It is easy to compute that

$$R_{kl} = R g_{kl}, \quad R = \frac{2R_{1212}}{g_{11}g_{22} - g_{12}^2}. \tag{10.41}$$

The following theorem establishes a relationship between the scalar and Gaussian curvatures in the two-dimensional case.

**Theorem 10.12.** *Let $M^2$ be a two-dimensional surface in $\mathbb{R}^3$ (with induced metric). Then its scalar curvature $R$ is twice the Gaussian curvature:*

$$R = 2K.$$

**Proof.** Take some point $P \in M^2$ and represent the surface in a neighborhood of this point as the graph of a function $z = f(x, y)$, assuming that the $z$-axis is directed along the normal to the surface at the point $P$. The metric on the surface is

$$g_{11} = 1 + f_x^2, \quad g_{12} = f_x f_y, \quad g_{22} = 1 + f_y^2,$$

and, since the $z$-axis is orthogonal to the surface at $P$, $\operatorname{grad} f(P) = 0$. We see that all the derivatives of the form $\frac{\partial g_{ij}}{\partial u^k}$ at $P$ vanish, hence the Christoffel symbols at this point also vanish, $\Gamma^k_{ij} = 0$. Therefore, at the point $P$ we have (see (10.33))

$$R^i_{lkj} = \frac{\partial \Gamma^i_{lj}}{\partial x^k} - \frac{\partial \Gamma^i_{lk}}{\partial x^j},$$
$$R_{ijkl} = \frac{1}{2}\left(\frac{\partial^2 g_{il}}{\partial u^j \partial u^k} + \frac{\partial^2 g_{jk}}{\partial u^i \partial u^l} - \frac{\partial^2 g_{ik}}{\partial u^j \partial u^l} - \frac{\partial^2 g_{jl}}{\partial u^i \partial u^k}\right),$$

where $u^1 = x$, $u^2 = y$. Substituting the expressions for the metric $g_{ij}$ into this formula, we obtain

$$R_{1212} = f_{xx} f_{yy} - f_{xy}^2 = K.$$

In this coordinate system, $\det(g_{ij}) = g = 1$, hence

$$R = \frac{2R_{1212}}{g} = 2R_{1212} = 2K.$$

Since the scalar curvature $R$ and the Gaussian curvature $K$ are scalars, the equality $R = 2K$ derived in some coordinate system is valid regardless of the choice of coordinates. Hence the theorem. □

Formula (10.41) can be rewritten as

$$K = \frac{R}{2} = \frac{\langle R(e_1, e_2)e_1, e_2\rangle}{\langle e_1, e_1\rangle\langle e_2, e_2\rangle - \langle e_1, e_2\rangle^2},$$

where $e_1, e_2$ is a basis in the tangent space. We see that this expression yields the Gaussian curvature, hence it does not depend on the choice of basis $e_1, e_2$.

In general, it follows from the symmetries of the curvature tensor that if, at some point of an $n$-dimensional Riemannian manifold, we consider the two-dimensional plane $\sigma$ spanned by tangent vectors $X$ and $Y$ in the tangent space, then the expression

$$K(\sigma) = \frac{\langle R(X, Y)X, Y\rangle}{\langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2}$$

depends only on the plane $\sigma$, but not on the choice of vectors $X$ and $Y$. It is called the *sectional curvature* of the Riemannian manifold along the plane (two-dimensional direction) $\sigma$.

As we see, in the two-dimensional case the sectional curvature coincides with the Gaussian curvature.

For a metric of the form

$$dx^2 + G(x, y)\, dy^2$$

the sectional curvature equals

$$K = -\frac{1}{\sqrt{G}}\frac{\partial^2\sqrt{G}}{\partial x^2}.$$

With the aid of this formula we can easily construct metrics of constant sectional curvature.

1) The metric of zero curvature corresponds to the Euclidean plane,

$$dl^2 = dx^2 + dy^2, \quad K = 0.$$

2) The metric of positive curvature $K$ corresponds to the sphere of radius $\frac{1}{\sqrt{K}}$,

$$dl^2 = dx^2 + \sin^2\left(\sqrt{K}x\right) dy^2.$$

3) The metric of negative curvature $K$ corresponds to the pseudosphere,

$$dl^2 = dx^2 + \sinh^2\left(\sqrt{-K}x\right) dy^2.$$

The examples of coordinates in which the metrics on the sphere and pseudosphere take this form were given in Chapter 4. These are spherical and pseudospherical coordinates.

2. THREE-DIMENSIONAL CASE: $\dim M = 3$. In $n$-dimensional space, consider all skew-symmetric tensors of type $(2,0)$. They form a linear space $V$ of dimension $\frac{n(n-1)}{2}$ generated by the tensors

$$e_A = e_i \wedge e_j, \quad A = [ij], \quad 1 \le i < j \le n.$$

It follows from the symmetries of the curvature tensor that the latter determines a quadratic form on the space of all such tensors in the tangent space at each point, by the rule

$$\langle e_i \wedge e_j, e_k \wedge e_l \rangle = R_{ijkl} = R_{AB}, \quad A = [ij], \quad B = [kl],$$

which by bilinearity yields the general formula

$$\langle \xi, \eta \rangle = R_{AB} \xi^A \eta^B.$$

The space is said to have a *positive curvature operator* if at each point of this space the quadratic form $R_{AB}$ is positive definite on the space of skew-symmetric tensors of type $(2,0)$.

For $n = 3$, the space $V$ of skew-symmetric tensors of type $(2,0)$ is three-dimensional and a quadratic form on such a space is specified by six parameters. The Ricci tensor also has six independent components, being a quadratic form on a three-dimensional vector space. It turns out that in the three-dimensional case the Ricci tensor completely determines the Riemann tensor by the formula

$$R_{ijkl} = R_{ik}g_{jl} + R_{jl}g_{ik} - R_{il}g_{jk} - R_{jk}g_{il} + \frac{R}{2}\left(g_{il}g_{jk} - g_{ik}g_{jl}\right). \tag{10.42}$$

**10.2.4. Tensor of conformal curvature.** In the 4-dimensional case the Ricci tensor has 10 independent components, the space of skew-symmetric tensors of type $(2,0)$ is 6-dimensional, and a quadratic form on such a space is determined by 21 parameters. In this case, as well as for higher dimensions, the Ricci tensor does not determine the Riemann tensor.

The following formula generalizes (10.42):

$$\begin{aligned} R_{ijkl} = {} & \frac{1}{n-2}\left(R_{ik}g_{jl} + R_{jl}g_{ik} - R_{il}g_{jk} - R_{jk}g_{il}\right) \\ & + \frac{R}{(n-1)(n-2)}\left(g_{il}g_{jk} - g_{ik}g_{jl}\right) + W_{ijkl}, \end{aligned} \tag{10.43}$$

where $n$ is the dimension of the space. The tensor $W_{ijkl}$ has zero trace,

$$W^i_{kli} = 0.$$

Hence it may be regarded as the skew-symmetric part of the Riemann tensor, i.e., the part that makes no contribution into the Ricci tensor. This tensor is called the *Weyl tensor* of the *tensor of conformal curvature.*

Let us explain the last term.

A positive definite (Riemannian) metric $g_{ij}\,dx^i\,dx^j$ is said to be *conformally flat* if in a neighborhood of any point it can be reduced by a change of coordinates to the form

$$g_{ij} = e^u \delta_{ij},$$

or, equivalently, if upon multiplying by the positive function $e^{-u}$, the metric reduces to the planar form, i.e., becomes Euclidean in some coordinates.

In 4.4.1 we have shown that in the two-dimensional case any metric is conformally flat. For higher dimensions this is not the case.

The *Schouten tensor* is the tensor

$$S_{kl} = \frac{1}{n-2}\Big(R_{kl} - \frac{R}{2(n-1)} g_{kl}\Big).$$

One can define a symmetric multiplication on the space of quadratic forms resulting in tensors of type $(0,4)$, by the formula

$$(f \circ g)_{ijkl} = \det \begin{pmatrix} f_{il} & g_{ik} \\ f_{jl} & g_{jk} \end{pmatrix} + \det \begin{pmatrix} f_{jk} & g_{jl} \\ f_{ik} & g_{il} \end{pmatrix}.$$

In terms of this multiplication the formula for the Weyl tensor simplifies to

$$\begin{aligned} W_{ijkl} &= R_{ijkl} - \frac{1}{n-2}(R \circ g)_{ijkl} + \frac{2R}{(n-1)(n-2)}(g \circ g)_{ijkl} \\ &= R_{ijkl} - (S \circ g)_{ijkl}. \end{aligned}$$

The tensor

$$T_{ijk} = (\nabla_i S)_{jk} - (\nabla_j S)_{ik}$$

is called the *Weyl–Schouten tensor.*

In the three-dimensional case the Weyl tensor always vanishes, which follows from (10.42). In general, one can verify by an easy calculation that the Weyl tensor always vanishes for a conformally flat metric. Hence it can be viewed as an obstruction to reducing the metric to the conformally Euclidean form $g_{ij} = e^u \delta_{ij}$ in a domain of the space (obviously, this can always be done at a single point). We state the following theorem without proof.

**Theorem 10.13.** 1. *A three-dimensional Riemannian manifold $M^3$ is conformally flat if and only if its Weyl–Schouten tensor vanishes:*

$$(\nabla_i S)_{jk} = (\nabla_j S)_{ik}, \quad i, j, k = 1, 2, 3.$$

2. *A Riemannian manifold $M^n$ of dimension $n \geq 4$ is conformally flat if and only if its Weyl tensor vanishes*:

$$W_{ijkl} = 0, \quad i, j, k, l = 1, \dots, n.$$

Thus we conclude that starting from dimension 4, the Riemann tensor cannot be recovered from the Ricci tensor.

However, the case of dimension 4 with indefinite metric $g_{ij}$ of signature $(+---)$ provides the setup where the Ricci tensor plays an important role in physics.

The basic conception of Einstein's general relativity theory is that the gravitational field is a metric $g_{ij}$ of signature $(+---)$ (at each point, it reduces, by a change of coordinates, to the Minkowski metric $c^2\,dt^2 - dx^2 - dy^2 - dz^2$). The material fields (i.e., all the fields except for the gravitational one) occur in Einstein's equations

$$R_{kl} - \frac{R}{2} g_{kl} = \frac{8\pi G}{c^4} T_{kl}$$

via the energy-momentum tensor $\frac{8\pi G}{c^4} T_{kl}$, where $G$ is the gravitational constant and $c$ is the velocity of light in vacuum. If there are no material fields, then Einstein's equations become

$$R_{kl} - \frac{R}{2} g_{kl} = 0.$$

We find the trace of the tensor in the left-hand side of this formula:

$$R^k_k - \frac{R}{2} \delta^k_k = R - 2R = -R = 0,$$

which implies that without material fields, Einstein's equations are equivalent to the condition that the Ricci tensor vanishes,

$$R_{kl} = 0.$$

Manifolds with zero Ricci tensor and the corresponding metrics (of any signature) are said to be *Ricci flat*. Therefore, the Ricci flat manifolds with indefinite metrics of signature $(+---)$ exactly correspond to solutions of Einstein's equations in vacuum.

In modern differential geometry, a manifold for which the Ricci tensor is proportional to the metric,

$$R_{kl} = \lambda g_{kl}, \quad \lambda = \text{const},$$

is called *Einstein's manifold* irrespective of the signature of the metric. A metric proportional to its Ricci tensor is called *Einstein's metric*.

**10.2.5. Tetrad formalism.** There is one more formalism for dealing with curvature, which was developed by physicists to study Einstein's equations. It was called tetrad to emphasize the four-dimensional nature of the physical space, although we will present it for a general dimension.

Let $g_{ij}$ be a metric of fixed signature in $n$-dimensional space $M^n$. In this space, we introduce a constant metric $h_{ij}$ of the same signature. At each point of $M^n$ the metric $g_{ij}$ can be reduced to the form $h_{ij}$ by choosing basis vectors $e_1, \dots, e_n$ in the tangent space at this point such that

$$\langle e_i, e_j \rangle = h_{ij}.$$

Furthermore, we can achieve locally a smooth dependence of these basis vectors on the point. Thus we obtain a collection of smooth vector fields $e_1, \dots, e_n$ satisfying the equalities $\langle e_i, e_j \rangle = h_{ij}$. For example, in the general relativity theory it is expedient to choose a *tetrad*, i.e., a quadruple of vector fields $e_0, \dots, e_3$ such that

$$\langle e_0, e_0 \rangle = \langle e_1, e_1 \rangle = 0, \quad \langle e_0, e_1 \rangle = 1, \quad \langle e_2, e_2 \rangle = \langle e_3, e_3 \rangle = -1,$$

with all other scalar products of the basis vectors being equal to zero. Here $e_0$ and $e_1$ are light vectors, while $e_2, e_3$ are space-like.

Note that if the space is not flat, then these vector fields do not commute. Indeed, otherwise we could find local coordinates for which these vector fields would be basis fields and then to reduce the metric to a constant form in the entire domain of the space.

Consider the pairwise commutators $[e_i, e_j]$ of the fields and expand them in the same basis:

$$[e_i, e_j] = c_{ij}^k e_k.$$

Define the Christoffel symbols by the usual formula,

$$\nabla_{e_j} e_i = \Gamma_{ij}^k e_k.$$

**Lemma 10.5.** *If the connection is symmetric, then*

$$c_{ij}^k = \Gamma_{ji}^k - \Gamma_{ij}^k. \tag{10.44}$$

**Proof.** The symmetry implies that for any vector field the torsion tensor is equal to zero, $T(\xi, \eta) = 0$ (Lemma 10.4). In particular, we have

$$T(e_i, e_j) = \nabla_{e_i} e_j - \nabla_{e_j} e_i - [e_i, e_j] = (\Gamma_{ji}^k - \Gamma_{ij}^k - c_{ij}^k) e_k = 0.$$

Hence the lemma. □

With the aid of this lemma one can easily find the Christoffel symbols of the symmetric connection compatible with the metric.

**Theorem 10.14.** *If the connection is symmetric and compatible with the metric, then*

$$\Gamma^i_{jk} = \frac{1}{2} h^{il}(c_{jlk} + c_{klj} - c_{ljk}),$$

*where* $c_{ijk} = h_{il}c^l_{jk}$.

**Proof.** Since the connection is compatible with the metric, we have

$$0 = \nabla_{e_k}\langle e_i, e_j\rangle = \langle \nabla_{e_k} e_i, e_j\rangle + \langle e_i, \nabla_{e_k} e_j\rangle = \Gamma^l_{ik}h_{lj} + \Gamma^l_{jk}h_{il}.$$

Denote $\Gamma_{i,jk} = h_{il}\Gamma^l_{jk}$ and write down the last relation for all cyclic permutations of $i, j, k$; we obtain a system of three equations:

$$\Gamma_{i,kj} + \Gamma_{k,ij} = \Gamma_{k,ji} + \Gamma_{j,ki} = \Gamma_{j,ik} + \Gamma_{i,jk} = 0.$$

This system is solved taking into account (10.44) to yield the formulas for the coefficients of the connection. □

Using the formulas for the Christoffel symbols thus obtained, the curvature tensor (and hence the Einstein equations) can be written in terms of the functions $c^k_{ij}$ and their first-order derivatives, which is often helpful for computations.

**10.2.6. The curvature of invariant metrics of Lie groups.** Consider left-invariant metrics on Lie groups specified from the outset by means of "tetrads".

Let $G$ be a Lie group, and let $\mathfrak{g}$ be its Lie algebra identified with the tangent space to $G$ at the identity. Take some basis $e_1, \dots, e_n$ in $\mathfrak{g}$ and define the scalar products of these vectors:

$$\langle e_i, e_j\rangle = g_{ij}.$$

These scalar products uniquely determine a metric on the entire group $G$ by the following rule. To each element $g \in G$ corresponds a diffeomorphism, which is the left translation

$$h \to gh.$$

It smoothly depends on $g$ and induces an isomorphism of tangent spaces $T_1G \to T_gG$. Denote by $L_X$ the left-invariant vector field on the group $G$ obtained by translations of the vector $X \in \mathfrak{g} = T_1G$.

The collection of vector fields $L_{e_1}, \dots, L_{e_n}$ determines bases in the tangent spaces at the points $g \in G$ (see 8.3.5). The formula

$$\langle L_{e_i}, L_{e_j}\rangle = g_{ij}, \quad i, j = 1, \dots, n, \quad g \in G,$$

uniquely specifies a metric on the group. This metric is *left-invariant*, i.e., satisfies the equality

$$\langle X, Y\rangle = \langle L_X, L_Y\rangle$$

for all $X, Y \in T_gG$, $g \in G$. It is obvious that all the left-invariant metrics can be specified in this way.

Theorem 10.14 implies that the curvature of these metrics is determined by the scalar product $g_{ij}$ and commutation relations in the Lie algebra $\mathfrak{g}$ of the group $G$:

$$[L_{e_i}, L_{e_j}] = L_{[e_i, e_j]} = c_{ij}^k L_{e_k}.$$

Consider the simplest case where the Lie algebra is endowed with a Killing metric, i.e., a metric for which all the operators $\operatorname{ad} X$ are skew-symmetric,

$$\langle \operatorname{ad} X(Y), Z\rangle = -\langle Y, \operatorname{ad} X(Z)\rangle,$$

where $X, Y, Z \in \mathfrak{g}$.

**Theorem 10.15.** *The connection*

$$\nabla_{L_X} L_Y = \frac{1}{2}\, L_{[X,Y]} = \frac{1}{2}\,[L_X, L_Y]$$

*is symmetric and compatible with the Killing metric on the group $G$.*

**Proof.** We write down the torsion tensor of this connection:

$$T(L_X, L_Y) = \nabla_{L_X} L_Y - \nabla_{L_Y} L_X - [L_X, L_Y] = \frac{1}{2}\, L_{[X,Y]} - \frac{1}{2}\, L_{[Y,X]} - L_{[X,Y]} = 0.$$

It suffices to verify the compatibility condition only for left-invariant vector fields that determine the basis in the tangent space at each point:

$$\nabla_{L_X}\langle L_Y, L_Z\rangle = \langle \nabla_{L_X} L_Y, L_Z\rangle + \langle L_Y, \nabla_{L_X} L_Z\rangle.$$

The equality $\langle L_Y, L_X\rangle = \text{const}$ implies that

$$\nabla_{L_X}\langle L_Y, L_Z\rangle = \partial_{L_X}\langle L_Y, L_Z\rangle = 0.$$

Since all the operators $\operatorname{ad} X : Y \to [X, Y]$ are skew-symmetric relative to the Killing metric, we have

$$\begin{aligned}\langle \nabla_{L_X} L_Y, L_Z\rangle + \langle L_Y, \nabla_{L_X} L_Z\rangle &= \frac{1}{2}\,(\langle L_{[X,Y]}, L_Z\rangle + \langle L_Y, L_{[X,Z]}\rangle)\\ &= \frac{1}{2}\,(\langle [X,Y], Z\rangle + \langle Y, [X,Z]\rangle) = 0.\end{aligned}$$

Hence the theorem. □

**Corollary 10.8.** *The curvature of a symmetric connection compatible with the Killing metric is equal to*

$$R(L_X, L_Y)L_Z = \frac{1}{4}\, L_{[[X,Y],Z]}, \quad \langle R(L_X, L_Y)L_Z, L_W\rangle = \frac{1}{4}\,\langle [X,Y], [Z,W]\rangle.$$

**Proof.** We have

$$\begin{aligned}R(L_X, L_Y)L_Z &= \nabla_{L_Y}\nabla_{L_X} L_Z - \nabla_{L_X}\nabla_{L_Y} L_Z + \nabla_{[L_X, L_Y]} L_Z\\ &= \frac{1}{4}\,[L_Y, [L_X, L_Z]] - \frac{1}{4}\,[L_X, [L_Y, L_Z]] + \frac{1}{2}\,[[L_X, L_Y], L_Z],\end{aligned}$$

and taking into account the Jacobi identity we obtain

$$R(L_X, L_Y)L_Z = \frac{1}{4}\,[[L_X, L_Y], L_Z].$$

For the Killing metric the operator ad $Z$ is skew-symmetric, hence

$$\langle R(L_X, L_Y)L_Z, L_W\rangle = \frac{1}{4}\,\langle [[X,Y],Z],W\rangle = \frac{1}{4}\,\langle [X,Y],[Z,W]\rangle.$$

The proof is completed. □

## 10.3. Geodesic lines

**10.3.1. Geodesic flow.** In spaces with connection there are analogs of straight lines called geodesics and defined again by analogy with Euclidean geometry.

A curve $x(t)$ is called a *geodesic* if its velocity vector is parallel along the curve:

$$\nabla_{\dot{x}}\dot{x} = 0. \tag{10.45}$$

By the definition of the covariant derivative along a curve (see (10.18)) equation (10.45) can be rewritten in coordinates as

$$(\nabla_{\dot{x}}\dot{x})^i = \ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k = 0, \quad i = 1,\dots,n.$$

The equation thus obtained,

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{jk}\frac{dx^j}{dt}\frac{dx^k}{dt} = 0, \quad i = 1,\dots,n, \tag{10.46}$$

is called the *geodesic equation*.

If $\Gamma^j_{ki} = 0$, the solutions to this equation are ordinary straight lines, in accordance with Euclidean geometry.

For an arbitrary connection, (10.46) is a system of differential equations of second order. In a neighborhood of a point $(x_0^1,\dots,x_0^n)$ there exists a unique solution to this equation with initial conditions

$$x^j|_{t=0} = x_0^j, \quad \left.\frac{dx^j}{dt}\right|_{t=0} = \dot{x}_0^j, \quad j = 1,\dots,n,$$

for any $x_0^j$ and $\dot{x}_0^j$ (by the theorem on the existence and uniqueness of solutions). Hence we obtain the following lemma.

**Lemma 10.6.** *Given a connection $(\Gamma^i_{jk})$, for any point $x$ and any vector $v$ at this point there exists a unique geodesic of this connection starting from $x$ with initial velocity vector $v$.*

REMARK. It is seen from (10.46) that the geodesics of a connection depend only on its "symmetric part" $\Gamma^i_{(jk)} = \Gamma^i_{jk} + \Gamma^i_{kj}$.

If the connection is symmetric and compatible with the metric (Riemannian or pseudo-Riemannian), then the flow (10.46) on the tangent bundle is referred to as the geodesic flow of this metric.

**Lemma 10.7.** *The quantity*

$$E = \langle \dot{x}, \dot{x} \rangle = g_{ij}\dot{x}^i\dot{x}^j$$

*is preserved along the geodesic flow of the metric $g_{ij}$.*

**Proof.** The lemma is obtained by a direct calculation: we have

$$\frac{d}{dt} g_{ij}\dot{x}^i\dot{x}^j = \nabla_{\dot{x}} g_{ij}\dot{x}^i\dot{x}^j = (\nabla_{\dot{x}} g_{ij}) + g_{ij}[(\nabla_{\dot{x}}\dot{x})^i\dot{x}^j + \dot{x}^i(\nabla_{\dot{x}}\dot{x})^j] = 0,$$

since $\nabla_{\dot{x}}\dot{x} = 0$ by definition of the geodesic flow, while the property that the connection is compatible with the metric implies that $\nabla_{\dot{x}} g_{ij} = 0$. □

A geodesic flow has the form of an ordinary differential equation on the tangent bundle $TM^n$ to the manifold:

$$\dot{x}^i = \xi^i, \quad \dot{v}^i = -\Gamma^i_{jk}(x)v^j v^k, \quad i = 1, \ldots, n.$$

Moreover, we will state the following simple lemma.

**Lemma 10.8.** *If $x(t)$ is a trajectory of the geodesic flow and $C$ is a nonzero constant, then the curve $y(t) = x(Ct)$ is also a trajectory of the geodesic flow.*

**Proof.** Indeed,

$$\nabla_{\dot{y}}\dot{y} = \nabla_{C\dot{x}}(C\dot{x}) = C^2\nabla_{\dot{x}}\dot{x} = 0.$$

Hence the lemma. □

Henceforth we restrict ourselves to geodesic flows of Riemannian metrics.

For the geodesic flow of a metric $g_{ij}$ the tangent bundle fibers into submanifolds of the form $\{E = \text{const}\}$ invariant with respect to the flow. Lemma 10.8 implies that the restrictions of the geodesic flow to various levels $\{E = C_1\}$ and $\{E = C_2\}$ are equivalent in the sense of trajectories, i.e., they have the same trajectories if we regard them as curves. Indeed, we have the one-to-one correspondence

$$x(t) \subset \{E = C_1\} \longleftrightarrow y(t) = x(\lambda t) \subset \{E = C_2\}, \quad \text{where } \lambda = \sqrt{\frac{C_2}{C_1}}.$$

For $E = 0$ each trajectory consists of a single point (which is not the case for a pseudo-Riemannian metric). For this reason, by the geodesic flow of a Riemannian metric is usually meant the restriction of the geodesic flow to the level hypersurface $\{E = |\dot{x}|^2 = 1\}$.

The hypersurface $LM^n = \{|\dot{x}| = 1\}$ consisting of unit tangent vectors to $M^n$ is called the *manifold of unit elements* on $M^n$. If the manifold $M^n$ is compact, then $LM^n$ is also compact. As the following theorem shows, in this case we obtain a stronger property than the local existence of a solution.

**Theorem 10.16.** *Let $\dot{x} = v(x)$ be an ordinary differential equation with smooth right-hand side $v(x)$ on a compact manifold $M^n$ without boundary.*

*Then for any point $x_0 \in M^n$ there exists a solution $x(t)$ to this equation with initial condition $x(0) = x_0$, which is defined for any time $t \in \mathbb{R}$ and is a smooth function of the initial condition.*

**Proof.** By the local theorem on the existence and uniqueness of the solution to an ordinary differential equation, for any point $x_0$ there is a neighborhood $U = U(x_0)$ of this point and a constant $T = T(U) > 0$ such that for any $x' \in U$ there exists a unique solution $x(t)$, $t \in [0,T]$, of the equation with initial data $x(0) = x'$. Such domains $U(x_0)$ form an open covering of $M^n$, and due to compactness of $M^n$, this covering contains a finite subcovering $U_i$, $i = 1, \ldots, N$. Set $T = \min_{U_i} T(U_i)$.

We will show that there exists a unique solution for any $t \in \mathbb{R}$ and any initial condition $x(0) = x_0$. Denote by $\varphi_t \colon M^n \to M^n$ the translation along the solution of the equation. These mappings are defined for $t \in [0,T]$. Take an arbitrary value $t = kT + t'$, $t' \in [0,T)$, $k \in \mathbb{Z}$, and set

$$x(t) = \varphi_{t'}(\varphi_T)^k(x_0).$$

It is obvious that the curve $x(t)$ is defined for all $t \geq 0$ and determines a unique solution to the equation. Due to smooth dependence of the mapping $\varphi_t$ on the initial data, the value $x(t)$ for each time $t \geq 0$ is also a smooth function of the initial data. Applying the same reasoning to the reversed flow $\dot{x} = -v(x)$, we come to similar conclusions for $t \leq 0$. □

Applying this theorem to the geodesic flow restricted to the manifold of linear elements on a compact manifold, we obtain the following result.

**Corollary 10.9.** *On a closed Riemannian manifold, each geodesic $x(t)$ can be extended unboundedly (it is defined for all $t \in \mathbb{R}$).*

Riemannian manifolds on which all geodesics are extended unboundedly are said to be *complete* or *geodesically complete*. Such manifolds possess an important property to be stated without proof.

**Theorem 10.17.** *On a complete Riemannian manifold any two points can be joined by a geodesic.*

By means of geodesics one can construct special coordinate systems in a Riemannian manifold that are useful for computations and applications.

Let $x \in M^n$ and let $v$ be a tangent vector at $x$. Draw a geodesic $x(t)$ starting from $x = x(0)$ with initial velocity vector $\dot{x}(0) = v$. If the point $x(1)$ is defined, denote it by $\exp_x(v)$. The mapping

$$\exp\colon T_x M^n \to M^n$$

is called the *exponential mapping*. Here we indicated the entire tangent space as the domain of its definition, which is true for complete manifolds. In general, it follows from the local existence of geodesics that this mapping is defined at least for vectors of small length $|v| < \varepsilon$.

**Lemma 10.9.** *There is a ball $B = \{|v| < \eta\}$ in the tangent space at a point $x \in M^n$ such that*

$$\exp_x\colon B \to M^n$$

*maps this ball diffeomorphically onto a neighborhood of $x \in M^n$.*

**Proof.** Since $\frac{\partial(\exp_x)^i}{\partial v^j} = \delta^i_j$, the Jacobian is invertible, and by the inverse function theorem, the mapping $\exp_x$ is invertible on the image of a sufficiently small ball $B = \{|v| < \eta\}$. □

One can introduce new coordinates in the neighborhood $U$ onto which the ball $B$ is mapped diffeomorphically, by letting

$$P = (y^1, \ldots, y^n) \quad \text{if} \quad P = \exp_x(y).$$

Such coordinates are called *geodesic*. In physics they are also called *inertial*, since they possess an important property to be stated in the following lemma.

**Lemma 10.10.** *Let $x^1, \ldots, x^n$ be geodesic coordinates in a neighborhood of a point $P$ constructed by the mapping $\exp_P$. Then at the point $P$,*

$$\Gamma^i_{jk} = 0, \quad i, j, k = 1, \ldots, n.$$

**Proof.** In geodesic coordinates the geodesics with initial point $P$ and initial velocity vector $v$ have the form $x(t) = vt$. Substituting this formula into the geodesic equation we obtain $\Gamma^i_{jk} v^j v^k = 0$. Since the direction of the vector $v$ at the point $P$ can be chosen arbitrarily, all the Christoffel symbols vanish. Hence the lemma. □

**10.3.2. Geodesic lines as shortest paths.** Coordinates $x^1, \ldots, x^n$ are called *semigeodesic* if the metric tensor in these coordinates has the form

$$g_{ij}\, dx^i\, dx^j = \sum_{1 \le i,j \le n-1} g_{ij}\, dx^i\, dx^j + (dx^n)^2. \tag{10.47}$$

For example, the spherical and pseudospherical coordinates are semigeodesic (see Sections 4.2 and 4.3).

Semigeodesic coordinates have the following property.

**Lemma 10.11.** *In semigeodesic coordinates $x^1, \ldots, x^n$ any "straight line" of the form $(x^1, \ldots, x^{n-1}) = \text{const}$ is a geodesic with natural parameter $t = x^n$.*

**Proof.** By the formulas for Christoffel symbols we have

$$\Gamma^i_{nn} = \frac{1}{2} g^{ij}\Big(2\frac{\partial g_{jn}}{\partial x^n} - \frac{\partial g_{nn}}{\partial x^j}\Big) = 0, \quad \Gamma^n_{ij} = -\frac{1}{2}\frac{\partial g_{ij}}{\partial x^n}.$$

The "straight lines" $(x^1, \ldots, x^{n-1}) = \text{const}$, $x^n = t$ obviously satisfy the geodesic equations:

$$\begin{aligned}\ddot{x}^i + \sum_{1\le j,k\le n-1} \Gamma^i_{jk}\dot{x}^j\dot{x}^k + 2\sum_{1\le j\le n-1} \Gamma^i_{jn}\dot{x}^i\dot{x}^n + \Gamma^i_{nn}\dot{x}^n\dot{x}^n \\ = \ddot{x}^i + \sum_{1\le j,k\le n-1} \Gamma^i_{jk}\dot{x}^j\dot{x}^k + 2\sum_{1\le j\le n-1} \Gamma^i_{jn}\dot{x}^j\dot{x}^n = 0.\end{aligned}$$

Hence the lemma. □

The following lemma will be needed for the proof of the existence of semigeodesic coordinates.

**Lemma 10.12.** *Let $x \in M^n$, and let $S_{x,\tau}$ be spheres of the form $\{|v| = \tau\}$ in the tangent space at the point $x$. Then the geodesics $\exp_x(tv)$ are orthogonal to the submanifolds $\exp_x(S_{x,\tau})$ (for small $\tau$ such that the exponential mapping is defined).*

**Proof.** Let $v(s)$ be a smooth curve on the sphere $S_{x,\tau}$. It suffices to show that the tangent vector fields $F_t = \partial F/\partial t$ and $F_s = \partial F/\partial s$ to the surface $F(s,t) = \exp_x(tv(s))$ are everywhere orthogonal.

First, we prove that

$$\frac{D}{\partial t}\langle F_t, F_s\rangle = 0.$$

Since the connection is compatible with metric, we have

$$\frac{D}{\partial t}\langle F_t, F_s\rangle = \Big\langle \frac{D}{\partial t}F_t, F_s\Big\rangle + \Big\langle F_t, \frac{D}{\partial t}F_s\Big\rangle = \Big\langle \frac{D}{\partial t}F_s, F_t\Big\rangle.$$

Here we have used the fact that for all fixed $s$ the curves $F(s,t)$ are geodesic and $DF_t/\partial t \equiv 0$. The following identity can be easily verified to hold for imbedded surfaces (e.g., by expanding both sides in a basis):

$$\frac{D}{\partial t}F_s = \frac{D}{\partial s}F_t.$$

Thus we obtain

$$\frac{D}{\partial t}\langle F_s, F_t\rangle = \Big\langle \frac{D}{\partial s}F_t, F_s\Big\rangle = \frac{1}{2}\frac{D}{\partial s}\langle F_t, F_t\rangle = \frac{1}{2}\frac{d}{ds}|v|^2 = 0,$$

since $|v|^2 = \tau = \text{const}$. But $F_s = 0$ for $t = 0$. Therefore, $\langle F_s, F_t\rangle = 0$ for $t = 0$, and hence for all $t$. □

**Corollary 10.10.** *For any geodesic $x(t)$ and any of its points $P$, there are semigeodesic coordinates $x^1, \dots, x^n$ in a neighborhood of $P$ in which the geodesic $x(t)$ is specified by the equations $x^1 = \dots = x^{n-1} = 0$, $x^n(t) = t$.*

**Proof.** On the geodesic $x(t)$ take a point $Q$ close enough to $P$ for the mapping $\exp_Q$ to be defined in the ball $\{|v| < 2\eta\}$, and let $\exp_Q w = P$, where $|w| = \eta$. We take $x^1, \dots, x^{n-1}$ to be the coordinates on the sphere $\{|v| = \eta\}$ in $T_Q M^n$ such that the point $w$ has the coordinates $x^1 = \dots = x^{n-1}$. Take for $x^n$ the natural parameter on geodesics going from the point $Q$. By Lemma 10.12 these coordinates satisfy (10.47), i.e., they are semigeodesic. All the other properties required for these coordinates are fulfilled by construction. □

Semigeodesic coordinates allow for a simple proof of the fact that locally, geodesics are the shortest paths.

**Theorem 10.18.** *Let $x(t)$ be a geodesic, and $P$ a point on it. Then for any point of $x(t)$ sufficiently close to $P$, the geodesic between $P$ and this point has the smallest length among all curves joining these points.*

*If a curve $x(t)$ joining two points $P$ and $Q$ has the shortest length among the curves joining these points, then this curve is a geodesic.*

**Proof.** In a neighborhood of $P$, take semigeodesic coordinates given by Corollary 10.10. Suppose that they are defined in a domain $U$. Denote by $\delta$ the distance from $P$ to the boundary of $U$. Let a point $Q$ lie on a geodesic $x(t)$ at a distance $\rho < \delta$ from $P$, and let $\gamma$ be a curve joining the points $P$ and $Q$. If it goes somewhere outside the domain $U$, then its length is greater than $\delta > \rho$. Suppose that the curve lies entirely in $U$. Then its length $L$ satisfies the following inequalities:

$$L = \int_\gamma \sqrt{\sum_{1 \le i,j \le n-1} g_{ij}\dot{x}^i\dot{x}^j + (\dot{x}^n)^2}\, dt \ge \int_\gamma \sqrt{(\dot{x}^n)^2}\, dt = \int_\gamma dx^n = \rho,$$

and the equality is attained only for $\dot{x}^1 = \dots = \dot{x}^{n-1} = 0$, i.e., for an interval of the geodesic $x(t)$.

Now we show that the second statement is a consequence of the first. Any two sufficiently close points on the shortest curve can be joined by a geodesic, which must coincide with a part of the shortest curve. Since this holds in a neighborhood of any point on the curve, the entire shortest curve is a geodesic. The proof is completed. □

Globally geodesics are not the shortest paths. For example, as was shown in Section 4.2, the geodesics on the unit sphere in $\mathbb{R}^3$ are great circles. Any

two nonantipodal points split the great circle into two geodesic curves, one of which is the shortest path between these points, whereas the other is not.

**10.3.3. The Gauss–Bonnet formula.** Let $M^2$ be an oriented two-dimensional Riemannian manifold, and let $r(t)$ be a curve on this manifold. Suppose that the curve is regular, i.e., $\dot{r} \neq 0$ everywhere. The notion of curvature of a planar curve is extended to curves in $M^2$ in the following way.

Introduce the natural parameter on the curve, which will again be denoted by $t$, $|\dot{r}| \equiv 1$, and take the Frenet frames on the curve, i.e., the orthonormal positively oriented frames $(\dot{r}, n)$ in the tangent spaces at each point of the curve. The *geodesic curvature* is the quantity

$$k_g = \langle \nabla_{\dot{r}} \dot{r}, n \rangle.$$

When $M^2$ is the Euclidean plane, the geodesic curvature coincides with the curvature of the planar curve. Since $|\dot{r}| \equiv 1$, it can be proved as in the planar case that $\nabla_{\dot{r}} \dot{r} \perp \dot{r}$. This implies that a curve has zero geodesic curvature if and only if this curve is a geodesic.

**Theorem 10.19** (Gauss–Bonnet formula). *Let $V \subset M^2$ be a closed domain homeomorphic to a disk, and let its boundary $\partial V$ consist of successively traversed regular curves $r_1, \dots, r_n$ (the initial point of the curve $r_{i+1}$ is the terminal point of $r_i(t)$, and we set $r_{n+1} = r_1$). Let $\alpha_i$, $i = 1, \dots, n$, be the angle (taken inside $V$) between the curves $r_i$ and $r_{i+1}$ at their common endpoint. Then*

$$\int_{\partial V} k_g \, dt = 2\pi - \sum_{i=1}^{n} (\pi - \alpha_i) - \int_V K \, d\sigma, \tag{10.48}$$

*where the integral over the boundary contour is taken counterclockwise, $t$ is the natural parameter on the curve $\partial V$, and $d\sigma = \sqrt{g_{11}g_{22} - g_{12}^2}\, du^1 \wedge du^2$ is the oriented area form on the surface.*

**Proof.** Assume that the entire domain $V$ lies in a neighborhood where semigeodesic coordinates $(x, y)$ are introduced, $ds^2 = dx^2 + G\, dy^2$. We leave it as an exercise for the reader to derive the following formulas by direct calculations:

$$\Gamma^1_{22} = -\frac{1}{2} G_x, \quad \Gamma^2_{12} = \Gamma^2_{21} = \frac{1}{2G} G_x, \quad \Gamma^2_{22} = \frac{1}{2G} G_y,$$

while all the other Christoffel symbols are equal to zero. Here we let $u^1 = x$, $u^2 = y$. Hence we obtain the formula for the sectional curvature:

$$K = -\frac{1}{\sqrt{G}} \frac{\partial^2 \sqrt{G}}{\partial x^2}. \tag{10.49}$$

Let $e_1, e_2$ be bases in the tangent spaces corresponding to the coordinates $x, y$. The normal to the curve has the form

$$n = \frac{1}{\sqrt{G}}(-G\dot{y}e_1 + \dot{x}e_2).$$

Substitute these expressions into the formula for the geodesic curvature to obtain

$$k_g = \sqrt{G}\Big(-\ddot{x}\dot{y} + \dot{x}\ddot{y} + \frac{1}{2}G_x\dot{y}^3 + \frac{1}{G}G_x\dot{x}^2\dot{y} + \frac{1}{2G}G_y\dot{x}\dot{y}^2\Big).$$

Recall that $|\dot{r}|^2 = \dot{x}^2 + G\dot{y}^2 = 1$, whence we derive that

$$\frac{d}{dt}\arctan\Big(\frac{\sqrt{G}\dot{y}}{\dot{x}}\Big) = \sqrt{G}\Big(-\ddot{x}\dot{y} + \dot{x}\ddot{y} + \frac{1}{2G}G_x\dot{x}^2\dot{y} + \frac{1}{2G}G_y\dot{x}\dot{y}^2\Big),$$
$$(\sqrt{G})_x\dot{y} = (\dot{x}^2 + G\dot{y}^2)\frac{G_x\dot{y}}{2\sqrt{G}} = \sqrt{G}\Big(\frac{1}{2}G_x\dot{y}^3 + \frac{1}{2G}G_x\dot{x}^2\dot{y}\Big).$$

Thus we see that $k_g\,dt$ can be written as

$$k_g\,dt = (\sqrt{G})_x\dot{y}\,dt + d\arctan\Big(\frac{\sqrt{G}\dot{y}}{\dot{x}}\Big).$$

By Green's formula (9.17) (which is a particular case of the general Stokes formula), the integral of $(\sqrt{G})_x\dot{y}\,dt$ over the boundary is equal to

$$\int_{\partial V}(\sqrt{G})_x\dot{y}\,dt = \int_{\partial V}(\sqrt{G})_x\,dy = \int_V(\sqrt{G})_{xx}\,dx\wedge dy.$$

Using here the expression (10.49) for the Gaussian curvature we obtain

$$\int_{\partial V}(\sqrt{G})_x\dot{y}\,dt = -\int_V K\,d\sigma.$$

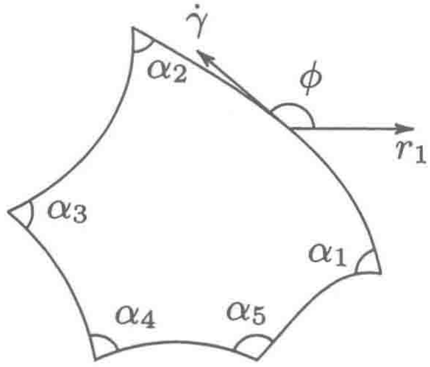


**Figure 10.2.** Geodesic polygon.

The angle $\arctan(\frac{\sqrt{G}\dot{y}}{\dot{x}})$ equals (up to $\pi$) the angle $\varphi$ between $\dot{r}$ and $e_1$. If $\partial V$ is a smooth curve, then $\int_{\partial V} d\varphi = 2\pi$. In the general case we have

$$\int_{\partial V} d\varphi = 2\pi - \sum_{i=1}^{n}(\pi - \alpha_i).$$

Finally, we obtain

$$\int_{\partial V} k_g\,dt = \int_{\partial V}\left(d\varphi + \left(\sqrt{G}\right)_x \dot{y}\,dt\right) = 2\pi - \sum_{i=1}^{n}(\pi - \alpha_i) - \int_V K\,d\sigma.$$

For a large domain $V$, the formula is proved as follows. Divide $V$ into small domains $V_i$ in which semigeodesic coordinates can be introduced, and sum up formulas (10.48) for $\int_{\partial V_i} k_g\,dt$ over these domains. The boundary contours of these domains lying inside $V$ will occur in the sum twice with opposite signs, hence $\int_{\partial V} k_g\,dt = \sum_i \int_{\partial V_i} k_g\,dt$. Summing up the right-hand sides of formulas (10.48) over $V_i$, we obtain the right-hand side of (10.48) for the domain $V$, and this completes the proof. □

A *geodesic triangle* on a surface $M^2$ is a domain homeomorphic to the interior of a triangle and bounded by three geodesic intervals. The sum of angles $\alpha_i$ between these intervals with vertices at their common endpoints is called, as in Euclidean geometry, the sum of angles of the triangle. For a geodesic triangle we have $k_g \equiv 0$ on the boundary, and (10.48) implies the following formula.

**Corollary 10.11.** *The sum of angles $\alpha_1, \alpha_2$, and $\alpha_3$ of a geodesic triangle $\Delta$ on a surface equals*

$$\alpha_1 + \alpha_2 + \alpha_3 = \pi + \int_\Delta K\,d\sigma.$$

*If the curvature of the surface is positive, then the sum of angles is greater than $\pi$, and if it is negative, the sum of angles is less than $\pi$.*

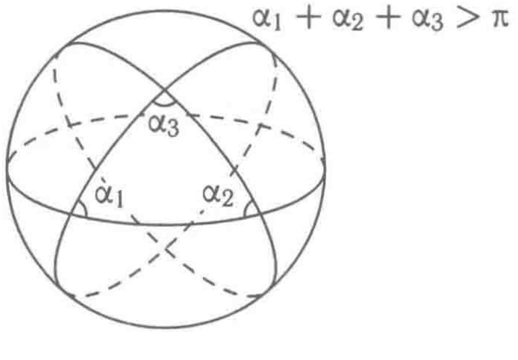


**Figure 10.3.** Geodesic triangle on a sphere.

## Exercises to Chapter 10

1. Prove that infinitesimal parallel translation of a vector $\xi^i$ by $\delta x^k$ transforms it by the formula

$$\xi^i \to \xi^i - \xi^j \Gamma^i_{jk} \delta x^k + o(|\delta x|).$$

2. Write down the deformation tensor in terms of covariant derivatives.

3. Prove that the trajectories of a point charge moving in the field of a magnetic pole determined by the equation

$$\ddot{r} = a \frac{[r, \dot{r}]}{|r|^3}, \quad a = \text{const},$$

are geodesics of a circular cone.

4. A two-dimensional metric is said to be a *Liouville metric* if it has the form $dl^2 = (f(u) + g(v))(du^2 + dv^2)$ in some coordinates. Prove that the level curves of the function

$$z(u, v) = \int \frac{du}{\sqrt{f(u) - a}} \pm \int \frac{dv}{\sqrt{g(v) + a}}$$

are geodesics of such a metric.

5. Prove that the meridians of a surface of revolution (the lines $\varphi =$ const) are geodesic lines, while a parallel of such a surface is geodesic if and only if the tangents to the meridians at its points are parallel to the axis of revolution.

6. Show that the geodesic lines of the metric $dl^2 = v(du^2 + dv^2)$ are parabolas on the plane $(u, v)$.

7. Prove that for a metric specified in polar coordinates by the formula $dl^2 = A\, dr^2 + r^2\, d\varphi^2$, the straight lines passing through the center are geodesics.

8. For a connected manifold, prove that:

a) a motion (isometry) which preserves a point and acts as the identity on the tangent space at this point is the identity;

b) the isometries of the manifold form a Lie group.

9. Prove that the isometries of a Killing metric on $\mathrm{SO}(n; \mathbb{R})$ that keep the unit element fixed are precisely the inner automorphisms $X \to AXA^{-1}$, where $X, A \in \mathrm{SO}(3, \mathbb{R})$.

10. Prove that a field $\xi = (\xi^i)$ is a Killing field if and only if

$$\nabla_i \xi_k + \nabla_k \xi_i = 0$$

for all $i, k$, where $\xi_i = g_{ik} \xi^k$.

11. Let $r(t)$ be a geodesic and $\xi$ a Killing field. Prove that the scalar product $\langle \dot{r}, \xi \rangle$ is constant along the geodesic.

12. Find all the Killing fields on the unit sphere in $\mathbb{R}^n$, $n \geq 3$.

13. Prove the following identities for a symmetric connection $\Gamma^i_{jk}$ compatible with a metric $g_{ij}$:

$$g^{kl}\Gamma^i_{kl} = -\frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^k}(\sqrt{|g|}g^{ik}), \quad \Gamma^i_{ki} = \frac{1}{2g}\frac{\partial g}{\partial x^k}.$$

14. For $n$-dimensional space with metric $(g_{ij})$, prove the formula

$$\oint_{\partial V} X^i\, dS_i = \int_V \nabla_i X^i \sqrt{|g|}\, dx^1 \wedge \cdots \wedge dx^n,$$

where

$$dS_i = \frac{1}{(n-1)!}\sqrt{|g|}\,\varepsilon_{i_1\ldots i_{n-1}i}\, dx^{i_1} \wedge \cdots \wedge dx^{i_{n-1}}.$$

15. Prove the following formula for three-dimensional manifolds:

$$R_{ijkl} = R_{ik}g_{jl} + R_{jl}g_{ik} - R_{il}g_{jk} - R_{jk}g_{il} + \frac{R}{2}(g_{il}g_{jk} - g_{ik}g_{jl}).$$

16. Let $\xi_\varepsilon$ be a vector obtained by parallel translation of a vector $\xi = (\xi^k)$ counterclockwise along the boundary of the square with side $\varepsilon$ spanned by the coordinate axes $x^i$, $x^j$. Prove that

$$\lim_{\varepsilon\to 0}\frac{\tilde{\xi}^k_\varepsilon - \xi^k}{\varepsilon^2} = R^k_{lji}\xi^l.$$

17. Derive from the Bianchi identity the following identity for the divergence of the Ricci tensor:

$$\nabla_l R^l_m = \frac{1}{2}\frac{\partial R}{\partial x^m}.$$

18. Let the sectional curvature $K$ at each point of the space be the same along all the two-dimensional directions. Prove that the curvature tensor has then the form

$$R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}),$$

and, using the Bianchi identity, deduce from this equality that the space has a constant sectional curvature.

19. Prove that the Weyl tensor $W_{ijk}$ is invariant under conformal transformations of the metric $g_{ik}(x) \to \lambda(x)g_{ik}(x)$.

20. Let $X_1, \ldots, X_n$ be vector fields that form a basis in the tangent space at each point of an $n$-dimensional manifold. Introduce the dual basis of 1-forms $\omega_1, \ldots, \omega_n$ specified by the condition $\omega_i(X_j) = \delta_{ij}$. Prove that

$$d\omega_k = -\frac{1}{2}c^k_{ij}\omega_i \wedge \omega_j, \quad \text{where } [X_i, X_j] = c^k_{ij}X_k$$

(here and in the next problem summation over repeated indices is assumed).

21. In the context of the previous problem, consider an affine connection specified by its values on the basis fields:

$$\nabla_{X_j} X_i = \Gamma^k_{ij} X_k.$$

Define 1-forms $\omega^i_j$ as

$$\omega^i_j = \Gamma^i_{jk}\omega^k.$$

These forms uniquely determine the connection $\nabla$. Let $T_{ij}$ and $R^i_{ljk}$ be the torsion and curvature tensors of this connection. Using formula (9.5), show that the following *Cartan's structural equations* hold:

$$d\omega_i = -\omega_{ij} \wedge \omega_j - \frac{1}{2} T^i_{jk}\omega_j \wedge \omega_k, \quad d\omega_{ij} = -\omega_{ik} \wedge \omega_{kj} + \frac{1}{2} R^i_{jlk}\omega_l \wedge \omega_k.$$

If the curvature equals zero, then the last equation becomes the *Maurer–Cartan equation*

$$d\Omega + \Omega \wedge \Omega = 0, \quad \Omega = (\omega_{ij}).$$

22. Let $x^1, \dots, x^n$ be a geodesic coordinate system with center at a point $P$. Prove that the following equalities hold at $P$:

$$\frac{\partial g_{ij}}{\partial x^k} = 0, \quad i, j, k = 1, \dots, n.$$

23. In the space $\mathbb{R}^2$ with coordinates $x, y$ take the half-plane $y > 0$ and define the metric on it by the formula

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

This gives us the *Lobachevsky plane.*

a) Prove that the sectional curvature of the Lobachevsky plane is everywhere equal to $K = -1$.

b) Prove that the geodesics of this metric in Euclidean coordinates $x, y$ are the rays orthogonal to the line $y = 0$ and the semicircles with centers on the line $y = 0$.

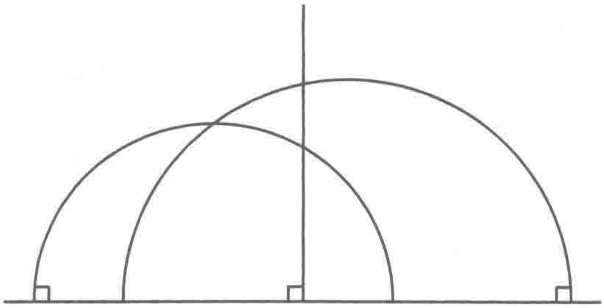

**Figure 10.4.** Geodesics on the Lobachevsky plane.

c) Prove that the Lobachevsky plane is geodesically complete and every two different points on it are joined by a unique geodesic segment, which is the shortest line between these points.

d) A geodesic will be said to be a "straight line" if it is unbounded in both directions (i.e., defined for all $t$). Prove that if a point $x$ does not belong to a "straight line" $\gamma$, then there are infinitely many different "straight lines" passing through $x$, which do not intersect $\gamma$ (Lobachevsky's axiom).

e) Prove that the action of the group $\mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$ on the Lobachevsky plane determined by the formula

$$z = x + iy \to \frac{az+b}{cz+d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R}),$$

preserves the lengths of all curves (is isometric).

f) Prove that for a geodesic triangle $ABC$ on the Lobachevsky plane with sides $BC$, $AC$, and $AB$ of length $a$, $b$, and $c$, respectively, and the angles at $A$, $B$, and $C$ equal to $\alpha$, $\beta$, and $\gamma$, the following equalities hold:

$$\cosh a = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\sin\gamma}, \quad \cosh b = \frac{\cos\beta + \cos\gamma\cos\alpha}{\sin\gamma\sin\alpha},$$

$$\cosh c = \frac{\cos\gamma + \cos\alpha\cos\beta}{\sin\alpha\sin\beta}.$$

g) Prove the following "law of sines" for the Lobachevsky plane:

$$\frac{\sinh a}{\sin\alpha} = \frac{\sinh b}{\sin\beta} = \frac{\sinh c}{\sin\gamma} = \frac{\sqrt{Q}}{\sin\alpha\sin\beta\sin\gamma},$$

where $Q = \cos^2\alpha + \cos^2\beta + \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma - 1$.

h) Find the length of a circle on the Lobachevsky plane as a function of its radius.

24. In the half-space $\{x^n > 0\}$ in $\mathbb{R}^n$, define the metric by the formula

$$g_{ij}\,dx^i\,dx^j = \frac{(dx^1)^2 + \cdots + (dx^n)^2}{(x^n)^2}$$

(the $n$-dimensional *Lobachevsky* or *hyperbolic space*). Prove that the sectional curvature of this metric at each point and in each two-dimensional direction is $K = -1$.

25. Prove that the geodesics of a Killing metric on a Lie group which pass through the identity are precisely the one-parameter subgroups $\exp(Xt)$.

26. Prove that if the covariant derivatives of a differential form $\omega$ on a Riemannian manifold vanish everywhere, $\nabla_\xi\omega \equiv 0$, then this form is closed, $d\omega = 0$.

Chapter 11

# Conformal and Complex Geometries

## 11.1. Conformal geometry

**11.1.1. Conformal transformations.** Let $g_{ij}$ and $g'_{ij}$ be two metrics defined in the same domain $U$ or, more generally, on the same manifold $M$. These metrics are said to be *conformally equivalent* if the angles between vectors in both metrics are the same. This is expressed simply as proportionality of the metrics:

$$g_{ij}(x) = \lambda(x) g'_{ij}(x), \quad \lambda(x) > 0.$$

Here $\lambda(x)$ is a scalar function depending on a point of the manifold, and the metrics are proportional with the same coefficient $\lambda(x)$ in any other coordinate system.

A mapping $f\colon M \to N$ of a manifold $M$ with coordinates $x^1, \dots, x^n$ and metric $g_{ij}\, dx^i\, dx^j$ into a manifold $N$ with coordinates $y^1, \dots, y^n$ and metric $g'_{ij}\, dy^i\, dy^j$ is said to be *conformal* if for any pair of vectors $\xi, \eta$ tangent to $M$ at the same point the angle between them is equal to the angle between their images $df(\xi)$ and $df(\eta)$ under the action of the differential. In tensor terms, it is obviously written as

$$g'_{ij}(y(x))\, dy^i\, dy^j = g'_{ij}(y(x)) \frac{\partial y^i}{\partial x^k} \frac{\partial y^j}{\partial x^l}\, dx^k\, dx^l = \lambda(x) g_{kl}(x)\, dx^k\, dx^l,$$

i.e.,

$$g'_{ij}(y(x)) \frac{\partial y^i}{\partial x^k} \frac{\partial y^j}{\partial x^l} = \lambda(x) g_{kl}(x)$$

for some positive scalar function $\lambda(x)$ on $M$.

EXAMPLE. STEREOGRAPHIC PROJECTION. Let $S^n$ be the unit sphere in $\mathbb{R}^{n+1}$ specified by the equation

$$(x^1)^2 + \cdots + (x^{n+1})^2 = 1.$$

We construct the stereographic projection of the sphere onto the space $\mathbb{R}^n$ specified by the equation $x^{n+1} = 0$. To this end we take the North Pole of the sphere $P_+ = (0, \ldots, 0, 1)$ and join it to each point $x \in S^n$ different from $P_+$ by a straight line $l_x$. Let $\pi(x)$ be the intersection point of $l_x$ with the plane $x^{n+1} = 0$. Thus we obtain the mapping

$$\pi \colon S^n \setminus \{P_+\} \to \mathbb{R}^n,$$

which maps the sphere with point $P_+$ removed onto the entire space $\mathbb{R}^n$. It is easily seen that this mapping is conformal. For $n = 2$ we already used it in 4.2.2 to construct complex coordinates on the two-dimensional sphere.

What are other conformal mappings known to us?

Obviously, for $n = 1$ any mapping is conformal.

For $n = 2$ any complex-analytic function $f \colon U \to \mathbb{C}$ specifies a conformal mapping of a small neighborhood $V$ of a nonsingular point $z_0 \in U$, $f_z(z_0) \neq 0$, onto the domain $f(V)$.

For $n \geq 3$ we indicate the following conformal mappings of the Euclidean space:

1) translations: $x \to x + a$, where $a \in \mathbb{R}^n$;

2) rotations: $x \to Ax$, where $A \in \mathrm{SO}(n)$;

3) *dilatations*: $x \to \lambda x$, where $\lambda$ is a nonzero constant;

4) *inversions* (with center at $x_0 \in \mathbb{R}^n$):

$$x \to x_0 + \frac{x - x_0}{|x - x_0|^2}.$$

Note that an inversion takes the point $x_0$ (the center of inversion) into the "point at infinity". The precise meaning of this is as follows. Each mapping $f$ of the form 1) to 4) can be extended to a conformal mapping of the sphere $S^n$ into itself by the rule:

$$y \in S^n \to \pi^{-1} f\, \pi(y).$$

Therefore, by a conformal mapping of $\mathbb{R}^n$ we mean a conformal mapping of the extended Euclidean space $\overline{\mathbb{R}^n} = S^n$ into itself.

For $n = 2$ we considered two models of the Lobachevsky plane $L^2$ (see 4.2.2):

1) Poincaré's model, which is the circle $|z| < 1$ on the complex plane with metric $\frac{4\,dz\,d\bar{z}}{(1-|z|^2)^2}$;

2) the upper half-plane $y > 0$ with metric $\frac{dx^2+dy^2}{y^2}$.

For $n \geq 3$, an analog of the upper half-plane with hyperbolic metric is the half-space $x^n > 0$ with metric

$$\frac{(dx^1)^2 + \cdots + (dx^n)^2}{(x^n)^2}, \quad x^n > 0,$$

while an analog of Poincaré's model is the ball $|x| < 1$ in $\mathbb{R}^n$ with metric $4\frac{|dx|^2}{(1-|x|^2)^2}$. These models of the $n$-dimensional Lobachevsky space (or *hyperbolic space*) $L^n$ are related by an isometry that continues to the boundary, on which this isometry is conformal. The sectional curvatures of these metrics are equal to $-1$ in any two-dimensional direction.

In Section 4.3 we used one more representation of the Lobachevsky plane as a pseudosphere $(x^0)^2 - \sum_{i=1}^n (x^i)^2 = 1$ with induced metric from $\mathbb{R}^{1,n}$ $(n = 2)$. A similar representation (as a sheet of a hyperboloid in $\mathbb{R}^{1,n}$) is valid for any hyperbolic space $L^n$, and one can prove, by means of this representation, that the group of orientation-preserving isometries of $L^n$ is the connected component of the group $\mathrm{SO}(1,n)$ that contains the identity. In this case the action of the group $\mathrm{SO}(1,n)$ on the hyperboloid is a restriction of the linear action of the group on $\mathbb{R}^{1,n}$.

The proof of all these facts for $n = 2$ given in Section 4.3 can be easily extended to an arbitrary $n$.

**Lemma 11.1.** *Translations, dilatations, rotations, and inversions of the space $\mathbb{R}^n$ regarded as the $n$-dimensional boundary plane of the $(n+1)$-dimensional half-space $L^{n+1}$ with hyperbolic metric are extendable to isometries of the space $L^{n+1}$.*

**Proof.** We will indicate these isometries explicitly. Let $x = (x', y)$, where $x' = (x^1, \ldots, x^n)$, and let the metric in $L^{n+1}$ be

$$ds^2 = \frac{\sum_{i=1}^n (dx^i)^2 + dy^2}{y^2}, \quad y > 0.$$

Then the isometries are given by the following formulas:

1) translation: $(x', y) \to (x' + a, y)$;

2) rotation: $(x', y) \to (Ax', y)$;

3) dilatation: $x \to \lambda x$;

4) inversion (with center at $x_0$): $x \to x_0 + \frac{x-x_0}{|x-x_0|^2}$. □

Any geodesic of the Lobachevsky metric (with a given direction of the curve) tends asymptotically to some point of the boundary. Let $\psi$ be an isometry of the Lobachevsky space $L^{n+1}$. If a geodesic $\gamma$ tends to a point

$P \in S^n = \overline{\mathbb{R}^{n-1}}$, and the geodesic $\psi(\gamma)$ to a point $Q \in S^n$, we set

$$\psi(P) = Q.$$

Thus with each isometry of the space $L^{n+1}$ we associate a mapping $\psi\colon S^n \to S^n$. All conformal transformations of the sphere $S^n$ indicated in Lemma 11.1 are obtained in this way.

Let us roughly calculate the number of parameters: the dimension of the group $\mathrm{SO}(1, n+1)$ is $1 + \cdots + (n+1) = (n+1)(n+2)/2$. The number of parameters specifying a group generated by conformal transformations of the form 1) to 4) is $n+n(n-1)/2+1+n = (n+1)(n+2)/2$, where the first and the last terms correspond to translations ($n$-dimensional translation vector) and inversions ($n$ coordinates of the center of inversion), respectively. We have compared the dimensions of the corresponding Lie algebras, and now it can be easily shown that in fact these groups coincide.

**Theorem 11.1.** *Each isometry of the $(n+1)$-dimensional Lobachevsky space $L^{n+1}$ induces a conformal mapping of the $n$-dimensional "boundary" sphere $S^n = \overline{\mathbb{R}^n}$, which mapping is a composition of translations, rotations, dilatations, and inversions. This correspondence is one-to-one.*

**11.1.2. Liouville's theorem on conformal mappings.** It turns out that for $n \geq 3$ all the conformal transformations of the sphere $S^n$ are generated by isometries of $L^{n+1}$ as described above. This is a consequence of the following theorem.

**Theorem 11.2** (Liouville). *Let $f\colon U \to V$ be a smooth conformal mapping of domains $U$ and $V$ in Euclidean space $\mathbb{R}^n$, where $n \geq 3$. Then $f$ is a composition of a motion, dilatation, and inversion.*

**Corollary 11.1.** *The group of conformal transformations of the sphere $S^n$ for $n \geq 3$ coincides with the group of isometries of the Lobachevsky space $L^{n+1}$.*

It can be proved by similar methods that the group of conformal transformations of the pseudo-Euclidean space $\mathbb{R}^{p,q}$ is $\mathrm{O}(p+1, q+1)$, $p + q \geq 3$. In particular, for the Minkowski space $\mathbb{R}^{1,3}$ we obtain $\mathrm{O}(2,4)$ as the group of conformal transformations.

The following lemma will be needed for the proof of the theorem.

**Lemma 11.2.** *Suppose that a metric $\lambda(x)((dx^1)^2+\cdots+(dx^n)^2)$ in a domain $U \subset \mathbb{R}^n$, $n \geq 3$, is Euclidean in some coordinates. Then either $\lambda \equiv \mathrm{const}$, or*

$$\lambda = \frac{1}{\rho^2}, \qquad \text{where} \quad \rho = \sum_{i=1}^{n} (ax^i + b_i)^2, \quad a \neq 0. \tag{11.1}$$

**Proof.** Since the metric is Euclidean, its Riemann tensor $R^i_{ljk}$ is identically equal to zero,

$$R^i_{lkj} = \frac{\partial \Gamma^i_{lj}}{\partial x^k} - \frac{\partial \Gamma^i_{lk}}{\partial x^j} + \Gamma^i_{mk}\Gamma^m_{lj} - \Gamma^i_{mj}\Gamma^m_{lk} = 0,$$

where the Christoffel symbols are calculated by the formula

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}\Big(\frac{\partial g_{lj}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\Big).$$

Substituting the metric

$$g_{ij} = \lambda\delta_{ij} = \frac{1}{\rho^2}\delta_{ij}$$

into these formulas, we obtain

$$\Gamma^i_{ii} = -\frac{\rho_{x^i}}{\rho}, \quad \Gamma^i_{ij} = \Gamma^i_{ji} = -\frac{\rho_{x^j}}{\rho}, \quad \Gamma^j_{ii} = \frac{\rho_{x^j}}{\rho}, \quad \Gamma^k_{ij} = 0,$$

where $i, j, k$ are pairwise different indices running over $1, \dots, n$. Here and in the sequel, a subscript such as $x^j$ means differentiation with respect to $x^j$. For the components of the Riemann tensor we have

$$R^k_{ikj} = \frac{\partial}{\partial x^j}\Big(\frac{\rho_{x^i}}{\rho}\Big) + \Big(\frac{\rho_{x^i}}{\rho}\Big)\Big(\frac{\rho_{x^j}}{\rho}\Big) = \frac{\rho_{x^i x^j}}{\rho} = 0,$$

where the indices $i, j, k$ are different and no summation over the repeated index $k$ is assumed (it is the derivation of this equality where the condition $n \geq 3$ is important). We have proved that

$$\rho_{x^i x^j} = 0, \quad 1 \leq i < j \leq n.$$

In a similar way, for $R^i_{iji}$, we obtain

$$R^i_{iji} = \frac{1}{\rho}(\rho_{x^i x^i} + \rho_{x^j x^j}) - \frac{1}{\rho^2}\sum_{k=1}^{n}(\rho_{x^k})^2 = 0,$$

where $i \neq j$. This implies the equalities

$$\rho(\rho_{x^i x^i} + \rho_{x^j x^j}) = (\rho_{x^1})^2 + \dots + (\rho_{x^n})^2, \quad 1 \leq i < j \leq n.$$

Since the right-hand side of the last equality does not depend on $i, j$, we have

$$\rho_{x^1 x^1} = \dots = \rho_{x^n x^n}.$$

These equalities, which hold in a domain of $\mathbb{R}^n$, imply that the function $\rho$ has the form

$$\rho(x^1, \dots, x^n) = a(x^1)^2 + \dots + a(x^n)^2 + b_1 x^1 + \dots + b_n x^n + b_0,$$

where

$$4a\rho = 2\rho\rho_{x^1 x^1} = (2ax^1 + b_1)^2 + \dots + (2ax^n + b_n)^2.$$

If $a = 0$, then $\rho = b_0 = \text{const}$. If $a \neq 0$, then $\rho^2 = \lambda^{-1}$ has the form (11.1). The proof is completed. □

**Proof of Theorem 11.2.** Let $f\colon U \to V$ be a conformal mapping of domains in $\mathbb{R}^n$, $n \geq 3$, with coordinates $x^1, \dots, x^n$ and $y^1, \dots, y^n$, respectively. Since this mapping is invertible in a small neighborhood of any point $y \in f(U)$ (by the inverse function theorem), we can take $x^1, \dots, x^n$ for the coordinates of the point $y = y(x^1, \dots, x^n)$ in this neighborhood. The Euclidean metric $(dy^1)^2 + \dots + (dy^n)^2$ in these coordinates becomes

$$\frac{\partial y^i}{\partial x^k}\frac{\partial y^j}{\partial x^l}\delta_{ij} = \lambda(x)\delta_{kl}$$

(due to the fact that the mapping is conformal). Since this metric is Euclidean, we can apply Lemma 11.2, which implies that either

1) $\lambda = c^2 = \text{const}$, or

2) $\lambda = \rho^{-2}$ with $\rho(x) = \sum_{i=1}^n (ax^i + b_i)^2$, $a \neq 0$.

Such conformal factors $\lambda(x)$ arise, respectively, when the mapping is either

1) the dilatation $y(x) = (cx^1, \dots, cx^n)$, or

2) the composition of the inversion centered at the point $\left(-\frac{b^1}{a}, \dots, -\frac{b^n}{a}\right)$ and the dilatation with coefficient $a^{-1}$.

Denote these conformal mappings by $g\colon \overline{\mathbb{R}^n} \to \overline{\mathbb{R}^n}$, and let $W$ be the domain $W = g(U)$.

We see that the mapping $fg^{-1} = h\colon W \to V$ is a motion, hence $f = hg$ is a composition of the motion $h$ and the mapping $g$, the latter being a composition of inversion and dilatation or simply a dilatation. Hence the theorem. □

**11.1.3. Lie algebra of a conformal group.** In 8.3.4 we realized Lie algebras by means of vector fields, or, equivalently, by means of first-order differential operators. Here we present such realizations for the Lie algebras of conformal groups.

Let us introduce the following fields written as first-order differential operators:

$$\Omega_{ij} = g_{ik}x^k\frac{\partial}{\partial x^j} - g_{jk}x^k\frac{\partial}{\partial x^i} \quad \text{(pseudorotation)},$$

$$P_i = \frac{\partial}{\partial x^i} \quad \text{(translation)},$$

$$D = x^i\frac{\partial}{\partial x^i} \quad \text{(dilatation)},$$

$$K_i = 2g_{ik}x^k x^j\frac{\partial}{\partial x^j} - g_{jk}x^j x^k\frac{\partial}{\partial x^i} \quad \text{(inversion)},$$

where $i, j = 1, \dots, n$.

For the Euclidean metric $g_{ij} = \delta_{ij}$, the transformation $\exp(t\Omega_{ij})$ specifies a rotation in the plane $(x^i, x^j)$. For a pseudo-Euclidean metric $g_{ij} = \lambda_i\delta_{ij}$,

$\lambda_i = \pm 1$, the transformation $\exp(t\Omega_{ij})$ specifies either a rotation (if $\lambda_i = \lambda_j$), or an elementary Lorentz transformation (if $\lambda_i = -\lambda_j$) in the plane $(x^i, x^j)$. The transformations $\exp\left(t\frac{\partial}{\partial x^i}\right)$ are translations along the $x^i$-axis, and the transformation $\exp\left(tx^i\frac{\partial}{\partial x^i}\right)$ is a dilatation $D(x) = tx$. The fields $K_i$ are nonlinear, but it can be proved that the group $\exp(tK_i)$ specifies dilatations with centers on the $x^i$-axis.

Any conformal transformation on $\mathbb{R}^{p,q}$ close to the identity is representable as $\exp(tA)$, where

$$A = \lambda^{ij}\Omega_{ij} + \mu^i P_i + \gamma D + \delta^i K_i.$$

The vector fields $\Omega_{ij}, P_i, D, K_i$, $i, j = 1, \dots, n$, form a Lie algebra $L$ with commutation relations

$$[\Omega_{ij}, \Omega_{kl}] = g_{ik}\Omega_{jl} - g_{jk}\Omega_{il} + g_{il}\Omega_{kj} - g_{jl}\Omega_{ki},$$
$$[\Omega_{ij}, P_k] = g_{ik}P_j - g_{jk}P_i, \quad [\Omega_{ij}, K_l] = g_{il}K_j - g_{jl}K_i,$$
$$[\Omega_{ij}, D] = [P_i, P_j] = [K_i, K_j] = 0,$$
$$[P_i, K_j] = 2(g_{ij}D + \Omega_{ij}), \quad [P_i, D] = P_i, \quad [K_i, D] = -K_i.$$

The following theorem can be verified directly.

**Theorem 11.3.** *If $g_{ab} = \delta_{ab}$ (Euclidean metric), then there is an isomorphism between the Lie algebra $L$ and the Lie algebra of the group $\mathrm{O}(n+1, 1)$,*

$$L = o(n+1, 1).$$

*It is established by the relations*

$$\begin{aligned} \Omega_{ij} &\to \Omega'_{\mu\nu}, \quad \mu = i = 1, \dots, n, \quad \nu = j = 1, \dots, n, \\ P_i &\to \Omega'_{i,n+1} - \Omega'_{i,n+2}, \\ K_i &\to \Omega'_{i,n+1} + \Omega'_{i,n+2}, \\ D &\to \Omega'_{n+1,n+2}, \end{aligned} \tag{11.2}$$

*where the vector fields $\Omega'_{\mu\nu}$ are generators of pseudorotations in the planes $(x^\mu, x^\nu)$ in $\mathbb{R}^{n+1,1}$.*

For $n = 2$ there are many local conformal transformations, and of particular interest among them is the subgroup of linear-fractional transformations of the sphere $S^2 \to S^2$,

$$z \to \frac{az + b}{cz + d}.$$

This subgroup is important for the study of motions of the Lobachevsky plane $L^2$ (see Section 4.3). It is isomorphic to $\mathrm{SL}(2, \mathbb{C})/\pm 1$ and is generated (after transition to the stereographic projection) by rotations, translations, dilatations, and inversions in $\mathbb{R}^2$. The isomorphism (11.2) specifies an isomorphism between the Lie algebra $\mathfrak{sl}(2, \mathbb{C})$ consisting of all zero-trace

complex $2 \times 2$ matrices and the Lie algebra of the Lorentz group SO(3,1) realized as the group of pseudorotations (of the form $\exp(tA)$). The latter isomorphism is called the semispinor representation of the Lorentz group SO(3,1) by complex $2 \times 2$ matrices (see Section 15.2).

## 11.2. Complex structures on manifolds

**11.2.1. Complex differential forms.** If a manifold is endowed with complex coordinates (see 5.1.7 for examples of complex manifolds), then it is natural to consider complex-valued differential forms on such spaces and to study them by methods of complex analysis.

Let $U$ be a $2n$-dimensional domain with complex coordinates $z^1, \dots, z^n$. Consider the corresponding real coordinates $x^1, \dots, x^n, y^1, \dots, y^n$, where $z^k = x^k + iy^k$, $k = 1, \dots, n$.

In the tangent space $T_x$ at any point $x \in U$, there is a natural basis of tangent vectors:

$$\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial y^1}, \dots, \frac{\partial}{\partial y^n}.$$

Over the field of complex numbers, these vectors generate a complexified tangent space $T_x^{\mathbb{C}}$, in which there is another natural basis over the field $\mathbb{C}$:

$$\frac{\partial}{\partial z^k} = \frac{1}{2}\Big(\frac{\partial}{\partial x^k} - i\frac{\partial}{\partial y^k}\Big), \quad \frac{\partial}{\partial \bar{z}^k} = \frac{1}{2}\Big(\frac{\partial}{\partial x^k} + i\frac{\partial}{\partial y^k}\Big), \quad k = 1, \dots, n.$$

To each of these bases corresponds the dual basis in the space of complex-valued 1-forms:

$$dx^1, \dots, dx^n, dy^1, \dots, dy^n,$$

and

$$dz^k = dx^k + i\,dy^k, \quad d\bar{z}^k = dx^k - i\,dy^k, \quad k = 1, \dots, n,$$

respectively.

Note that multiplication of complex-valued forms by the imaginary unit $i = \sqrt{-1}$ corresponds to the linear mapping on the space of real forms:

$$dz^k = dx^k + i\,dy^k \to i\,dz^k = i\,dx^k - dy^k : \ J(dx^k) = -dy^k, \ J(dy^k) = dx^k,$$

and this mapping satisfies the equality

$$J^2 = -1.$$

Since in arbitrary coordinates the transformation $J$ is determined at each point by a matrix $J = (J^r_s)$, it specifies also an automorphism of tangent spaces

$$(J\xi)^r = J^r_s \xi^s, \quad J^2 = -1. \tag{11.3}$$

The operator $J$ (which is a field of tensors of type $(1,1)$) is said to specify the complex structure on the manifold.

Any complex-valued $p$-form $\omega$ is representable as a linear combination of forms

$$dz^{i_1} \wedge \cdots \wedge dz^{i_r} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_s}, \quad r+s=p,$$

i.e., the following decomposition holds:

$$\omega = \sum_{\substack{i_1<\cdots<i_r \\ j_1<\cdots<j_s}} T_{i_1\dots i_r\bar{j}_1\dots\bar{j}_s}\, dz^{i_1} \wedge \cdots \wedge dz^{i_r} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_s}, \tag{11.4}$$

where the components of the tensor $T_{i_1\dots i_r\bar{j}_1\dots\bar{j}_s}$ are skew-symmetric separately in the subscripts $i_1,\dots,i_r$ and $j_1,\dots,j_s$.

Here and after, when writing coordinates of tensors, we mark by a bar the indices related to the coordinates $\bar{z}^k$.

We see that each $p$-form $\omega$ admits a natural decomposition into a sum of forms $\omega_{r,s}$, which are linear combinations of products of $r$ forms of type $dz^j$ and $s$ forms of type $d\bar{z}^k$:

$$\omega = \omega_{p,0} + \omega_{p-1,1} + \cdots + \omega_{1,p-1} + \omega_{0,p}. \tag{11.5}$$

The form $\omega_{r,s}$ is said to be a *form of type* $(r,s)$.

Since under holomorphic changes of coordinates $w^j = w^j(z)$, $\bar{w}^j = \overline{w^j(z)}$, the forms $dz^k$ go into linear combinations of forms $dw^j$, and the forms $d\bar{z}^k$ go into linear combinations of forms $d\bar{w}^j$, the decomposition (11.5) is independent of the choice of complex coordinates on the manifold. Therefore, this decomposition generates a decomposition of the exterior differentiation operator.

**Lemma 11.3.** *The operator $d$ of exterior differentiation decomposes into the sum*

$$d = d' + d''.$$

*If $\omega$ is a form of type $(r,s)$, then $d'\omega$ is a form of type $(r+1,s)$ and $d''\omega$ is a form of type $(r,s+1)$. This decomposition is invariant with respect to complex-analytic changes of coordinates.*

**Proof.** If the form $\omega$ is given by (11.4), then we set

$$d'\omega = \sum_k \sum_{\substack{i_1<\cdots<i_r \\ j_1<\cdots<j_s}} \frac{\partial T_{i_1\dots i_r\bar{j}_1\dots\bar{j}_s}}{\partial z^k}\, dz^k \wedge dz^{i_1} \wedge \cdots \wedge dz^{i_r} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_s},$$

$$d''\omega = \sum_l \sum_{\substack{i_1<\cdots<i_r \\ j_1<\cdots<j_s}} \frac{\partial T_{i_1\dots i_r\bar{j}_1\dots\bar{j}_s}}{\partial \bar{z}^l}\, d\bar{z}^l \wedge dz^{i_1} \wedge \cdots \wedge dz^{i_r} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_s}.$$

We see that $d\omega = d'\omega + d''\omega$, hence we obtain a decomposition of the form $d\omega$ into the sum of an $(r+1,s)$-form $d'\omega$ and an $(r,s+1)$-form $d''\omega$. Uniqueness

of this decomposition implies that the decomposition $d = d' + d''$ is invariant under complex-analytic changes of coordinates. Hence the lemma. □

Since
$$d^2 = (d' + d'')^2 = d'^2 + d''^2 + (d'd'' + d''d') = 0,$$
and for an $(r,s)$-form $\omega$ the forms $d'^2\omega$, $d''^2\omega$, and $(d'd'' + d''d')\omega$ are of the types $(r+2,s)$, $(r,s+2)$, and $(r+1,s+1)$, respectively, we see that all these terms in the decomposition for $d^2\omega$ vanish, and so we obtain the following result.

**Corollary 11.2.** *The operators $d'$ and $d''$ satisfy the relations*

$$(d')^2 = (d'')^2 = 0, \quad d''d' = -d'd''.$$

Thus we see that on the space $C(M)$ of smooth forms on a complex manifold $M$, one can introduce two more differentials
$$d' \colon C^p \to C^{p+1}, \quad d'' \colon C^p \to C^{p+1}$$
and define the cohomology groups for the pair $(C,d)$ as was done in 9.3.1.

We restrict our attention to the differential $d''$. The forms of type $(r,0)$ that are closed relative to this differential, i.e., satisfy the relation
$$d''\omega = 0,$$
are called *holomorphic* forms.

In the case $r = 0$, the holomorphic $(0,0)$-forms are exactly the complex-analytic functions on the manifold $M$. In the general case, these are $(r,0)$-forms
$$\omega = \sum_{i_1<\dots<i_r} T_{i_1\dots i_r}\, dz^1 \wedge \dots \wedge dz^{i_r},$$
with coefficients $T_{i_1\dots i_r}$ that are complex-analytic functions.

The cohomology groups
$$H^p_{\bar\partial}(M) = \frac{\operatorname{Ker} d'' \colon C^p(M) \to C^{p+1}(M)}{\operatorname{Im} d'' \colon C^{p-1}(M) \to C^p(M)}$$
are called the *Dolbeault cohomology.*

These groups have one more grading related to the presence of the complex structure:
$$H^p_{\bar\partial}(M) = \sum_{r+s=p} H^{r,s}_{\bar\partial}(M),$$
where
$$H^{r,s}_{\bar\partial}(M) = \frac{\operatorname{Ker} d'' \colon C^{r,s}(M) \to C^{r,s+1}(M)}{\operatorname{Im} d'' \colon C^{r,s-1}(M) \to C^{r,s}(M)}$$
and $C^{r,s}(M)$ is the space of $(r,s)$-forms on the manifold $M$.

Unlike the de Rham cohomology, the Dolbeault cohomology groups may be quite large even for simple manifolds. For example, in the case of $\mathbb{C}^n$, the group $H^0_{\bar\partial}(\mathbb{C}^n)$ is the space of all holomorphic functions on $\mathbb{C}^n$.

**11.2.2. Kähler metrics.** In the complexified tangent space at a point, each vector $\xi$ can be expanded in the basis $\partial/\partial z^k$, $\partial/\partial \bar z^l$, $k, l = 1, \dots, n$:

$$\xi = \xi^k \frac{\partial}{\partial z^k} + \xi^{\bar l} \frac{\partial}{\partial \bar z^l}.$$

The real tangent vectors are singled out by the condition

$$\xi^{\bar k} = \overline{\xi^k},$$

which shows that any real tangent vector is uniquely determined by its complex coordinates $\xi^1, \dots, \xi^n$.

A Hermitian metric in a domain $U \subset M$ with complex coordinates $z^1, \dots, z^n$ is specified by the functions $g_{j\bar k}$ that determine the Hermitian scalar product of real tangent vectors by the rule

$$\langle \xi, \eta \rangle = g_{j\bar k} \xi^j \eta^{\bar k} = g_{j\bar k} \xi^j \overline{\eta^k}.$$

The Hermitian condition

$$\langle \xi, \eta \rangle = \overline{\langle \eta, \xi \rangle} \tag{11.6}$$

implies the following condition on the functions $g_{j\bar k}$:

$$g_{j\bar k} = \overline{g_{k\bar j}}.$$

Henceforth we will consider only positive definite scalar products. This means that

$$\langle \xi, \xi \rangle = g_{j\bar k} \xi^j \xi^{\bar k} > 0 \quad \text{for} \quad \xi \neq 0.$$

For a Hermitian scalar product to be invariant with respect to complex-analytic changes of variables, we must require that the functions $g_{j\bar k}$ constitute a tensor and transform under the changes $\tilde z^j = \tilde z^j(z^1, \dots, z^n)$ by the rule

$$\tilde g_{j\bar k} \frac{\partial \tilde z^j}{\partial z^l} \overline{\frac{\partial \tilde z^k}{\partial z^m}} = g_{l\bar m}.$$

Decompose the Hermitian scalar product into the real and imaginary parts:

$$\langle \xi, \eta \rangle = \operatorname{Re}\langle \xi, \eta \rangle + i \operatorname{Im}\langle \xi, \eta \rangle.$$

The Hermitian condition (11.6) implies that

$$\langle \eta, \xi \rangle = \operatorname{Re}\langle \eta, \xi \rangle + i \operatorname{Im}\langle \eta, \xi \rangle = \operatorname{Re}\langle \xi, \eta \rangle - i \operatorname{Im}\langle \xi, \eta \rangle = \overline{\langle \xi, \eta \rangle}.$$

Hence we make two conclusions.

1) The real part of a Hermitian scalar product specifies a Riemannian metric $(\xi, \eta) = \operatorname{Re}\langle \xi, \eta \rangle$ on the tangent space.

2) The imaginary part of a Hermitian scalar product is a skew-symmetric 2-form $\mathrm{Im}\langle\xi,\eta\rangle = -\Omega(\xi,\eta)$ on tangent vectors.

It is easily verified that the form $\Omega$ can be written as

$$\Omega = \frac{i}{2}\, g_{j\bar{k}}\, dz^j \wedge d\bar{z}^k. \tag{11.7}$$

One can always find coordinates in which the Hermitian metric at a given point is reduced to the diagonal form: $g_{j\bar{k}} = \delta_{jk}$. In this case the Riemannian metric is

$$ds^2 = \sum_{k=1}^{n} |dz^k|^2 = \sum_{k=1}^{n}((dx^k)^2 + (dy^k)^2) \tag{11.8}$$

and the form $\Omega$ equals

$$\Omega = \frac{i}{2}\sum_{k=1}^{n} dz^k \wedge d\bar{z}^k = \sum_{k=1}^{n} dx^k \wedge dy^k, \tag{11.9}$$

where $z^k = x^k + iy^k$, $k = 1,\ldots,n$.

It follows from (11.7) that the imaginary part of a Hermitian product specifies a 2-form on the entire manifold. Formulas (11.8) and (11.9) imply that in these coordinates

$$(\xi,\eta) = \Omega(\xi, J\eta), \tag{11.10}$$

where $J$ is the complex structure operator (11.3). Since this relation is of tensor nature, it holds in any coordinates.

A Hermitian metric $g_{j\bar{k}}$ on a complex manifold $M$ is called a *Kähler metric* if the form $\Omega$ as in (11.7) is closed:

$$d\Omega = 0.$$

When speaking about the curvature of a Kähler metric, we will mean the curvature of the Riemannian metric $\mathrm{Re}\langle\xi,\eta\rangle$. Naturally, a complex manifold $M$ with a Kähler metric is called a *Kähler manifold*, and the corresponding form $\Omega$, the *Kähler form*.

EXAMPLE 1. Let $M^2$ be a two-dimensional Riemannian manifold. Then one can introduce on it a conformal parameter $z = x + iy$ such that the metric becomes $g\, dz\, d\bar{z}$ (see 4.4.1). The corresponding 2-form $\Omega$ is written as $\Omega = \frac{i}{2} g\, dz \wedge d\bar{z}$. It is closed, as is any 2-form on a two-dimensional manifold. Therefore, any Riemannian metric on a surface determines a Kähler metric.

**Theorem 11.4.** *Let $g\, dz\, d\bar{z}$ be a conformally Euclidean metric on a two-dimensional manifold. Then the $(1,1)$-form*

$$\omega = -id'd'' \log g$$

*is well defined and satisfies the equality*

$$\omega = K\Omega, \tag{11.11}$$

*where $K$ is the Gaussian curvature of the surface (or the sectional curvature of the two-dimensional manifold) and $\Omega = \frac{i}{2}\, g\, dz \wedge d\bar{z}$ is the Kähler form.*

**Proof.** Note that $\log g$ is not a $(0,0)$-form, and it transforms under a change $\tilde{z} = f(z)$ of the conformal parameter by the rule

$$\log g \to \log \tilde{g} = \log\Big(g \frac{\partial f}{\partial z}\overline{\frac{\partial f}{\partial z}}\Big) = \log g + \log F + \log \bar{F},$$

where $F = \partial f/\partial z$ is the complex Jacobian of the parameter change. This change is holomorphic, i.e., $\frac{\partial F}{\partial \bar{z}} = 0$. Therefore, $d''F = d'\bar{F} = 0$, and taking into account the relation $d''d' = -d'd''$, we obtain

$$d'd''(\log g + \log F + \log \bar{F}) = d'd'' \log g + d'd'' \log F - d''d'\bar{F} = d'd'' \log g.$$

Therefore $\omega$ is well defined as $(1,1)$-form.

Now to complete the proof it remains to note that in conformally Euclidean coordinates the Gaussian curvature is given by the formula

$$K = -\frac{2}{g}\,\frac{\partial^2 \log g}{\partial z \partial \bar{z}}$$

(see 4.4.2) and

$$d'd'' \log g = \frac{\partial^2}{\partial z \partial \bar{z}} \log g\, dz \wedge d\bar{z}.$$

Hence the theorem. □

EXAMPLE 2. Let $M = \mathbb{C}^n/\Lambda$ be a complex torus, obtained as the quotient space of $\mathbb{C}^n$ by the action of the translation group $\Lambda$ with $2n$ generators. We choose the planar metric on the torus:

$$ds^2 = \sum_k |dz^k|^2,$$

where $z^1, \dots, z^n$ are linear coordinates defined modulo the lattice $\Lambda \subset \mathbb{C}^n$.

The following theorem shows that locally Kähler metrics look like a planar Kähler metric up to infinitesimal terms of second order.

**Theorem 11.5.** *Let $M$ be a manifold of complex dimension $n$ with Hermitian metric $\langle \xi, \eta \rangle$, let $\nabla$ be a symmetric connection compatible with the Riemannian metric $(\xi, \eta) = \mathrm{Re}\langle \xi, \eta \rangle$, and let $J$ be the complex-structure operator* (11.3).

*Then $M$ is a Kähler manifold if and only if at least one of the following three conditions is fulfilled:*

1) *in a neighborhood of each point $P \in M$ one can introduce complex coordinates $z^1, \dots, z^n$ such that the Hermitian metric at the point $P$ becomes*

$$g_{j\bar{k}} = \delta_{jk} + [2], \tag{11.12}$$

*where* [2] *are the second-order terms in the Taylor expansion of the metric;*

2) $\nabla_\xi \Omega \equiv 0$;

3) $\nabla_\xi J \equiv 0$.

REMARK. Property 2) means that parallel translations are not only orthogonal (preserve the Riemannian metric), but also unitary (preserve the Hermitian structure, i.e., both the metric and the form $\Omega$).

**Proof of Theorem 11.5.** 1. If the metric reduces to the form (11.12) in a neighborhood of a point $P$, this immediately implies that the differential of the form $\Omega = \frac{i}{2} g_{j\bar{k}}\, dz^j \wedge d\bar{z}^k$ vanishes at this point, $d\Omega(P) = 0$.

Assume now that $g_{j\bar{k}}$ is a Kähler metric and write down its Taylor expansion about the point $P$:

$$g_{j\bar{k}} = \delta_{jk} + a_{j\bar{k}l} z^l + a_{j\bar{k}\bar{l}} \bar{z}^l + [2],$$

where $z^j(P) = 0$, $j = 1, \dots, n$. Since the metric is Hermitian, we have

$$a_{k\bar{j}l} = \overline{a_{j\bar{k}\bar{l}}}, \quad a_{k\bar{j}\bar{l}} = \overline{a_{j\bar{k}l}}.$$

The condition $d\Omega = 0$ implies that

$$a_{j\bar{k}l} = a_{l\bar{k}j}, \quad a_{j\bar{k}\bar{l}} = a_{j\bar{l}\bar{k}}.$$

We leave it as an exercise to prove by a direct calculation that a change of coordinates

$$w^k = z^k + \frac{1}{2} b^k_{jl} z^j z^l, \quad b^k_{jl} = b^k_{lj} = a_{j\bar{k}l}, \quad k = 1, \dots, n,$$

reduces the Hermitian metric at the point $P$ to the form (11.12).

2. Now we will prove that conditions 1) and 2) are equivalent. Let a Hermitian metric be of the form (11.12). Then at the point $P$ all the Christoffel symbols $\Gamma^j_{kl}$ of the Riemannian metric $(\xi, \eta) = \operatorname{Re} g_{j\bar{k}} \xi^j \bar{\eta}^k$ vanish. Now formula (11.12) implies that

$$\nabla_\xi \Omega = \partial_\xi \Omega = 0,$$

where $\partial_\xi$ is the directional derivative in the direction of the vector $\xi$. Since the point $P$ and the vector $\xi \in T_P M$ were chosen arbitrarily, we conclude that $\nabla_\xi \Omega \equiv 0$ everywhere.

Assume now that condition 2) holds. Take the geodesic coordinates $x^1, \dots, x^n, y^1, \dots, y^n$ with center at the point $P$. Since again all the Christoffel symbols $\Gamma^j_{kl}$ at the point $P$ vanish, we obtain that at this point

$$\nabla_\xi \Omega = \partial_\xi \Omega.$$

By assumption, we have $\nabla_\xi \Omega \equiv 0$, which implies that

$$d\Omega(P) = 0.$$

Since the point $P$ is arbitrary, $d\Omega \equiv 0$, which, as was shown before, is equivalent to condition 1).

3. The formula $(\xi, \eta) = \Omega(\xi, J\eta)$ (see (11.10)) immediately implies the equivalence of conditions 2) and 3). Indeed, if $h_{kl}$ is the Riemannian metric, then $\Omega_{jl} = h_{jl} J^l_k$. It remains to apply the Leibniz rule for covariant differentiation

$$(\nabla_\xi \Omega)_{jl} = (\nabla_\xi h)_{jk} J^k_l + h_{jk} (\nabla_\xi J)^k_l$$

and remember that the connection is compatible with the metric, $\nabla_\xi h_{jl} \equiv 0$. The proof is completed. □

The fact that condition 1) of Theorem 11.5 is equivalent to the property that the manifold is Kählerian, implies the following statement.

**Corollary 11.3.** *An identity on a Kähler manifold that involves only the metric and its first-order derivatives is valid if and only if it holds for the planar metric $g_{j\bar{k}} = \delta_{jk}$ on $\mathbb{C}^n$.*

**11.2.3. Topology of Kähler manifolds.** First of all we state a theorem that will enable us to construct new examples of Kähler manifolds from previously known.

**Theorem 11.6.** 1. *If $M_1$ and $M_2$ are Kähler manifolds with Kähler metrics $g_1$ and $g_2$, then $M = M_1 \times M_2$ with metric $g = g_1 + g_2$ is a Kähler manifold.*

2. *Let $M$ be a Kähler manifold with metric $g$, and let $\varphi\colon N \to M$ be a complex-analytic imbedding of a manifold $N$ into $M$. Then the manifold $N$ with induced metric is also a Kähler manifold.*

**Proof.** Assertion 1 is obvious. To prove assertion 2, note that if $\Omega$ is a Kähler form on $M$, then the imaginary part of the induced metric on $N$ is the form $\varphi^*\Omega$. Since the differential commutes with the operation of restriction of the tensor $\varphi^*$, we obtain

$$d(\varphi^*\Omega) = \varphi^*(d\Omega) = 0.$$

Therefore, the induced metric on the submanifold $N$ is also Kählerian. Hence the theorem. □

Now we give some important examples of Kähler manifolds.

Consider the complex projective space $\mathbb{C}P^n$ with homogeneous coordinates $(z^1 : \cdots : z^{n+1})$. It can be represented as the quotient space of the sphere $S^{2n+1} \subset \mathbb{C}^{n+1}$ (specified by the equation $\sum_k |z^k|^2 = 1$) by the action of the group $S^1$: $(z^1, \ldots, z^{n+1}) \to (e^{i\varphi}z^1, \ldots, e^{i\varphi}z^{n+1})$, $\varphi \in \mathbb{R}$. The projection $\pi\colon S^{2n+1} \to \mathbb{C}P^n$ has the maximal rank, equal to the dimension of the space $\mathbb{C}P^n$. Therefore, if $\pi^*\omega = 0$ for some form $\omega$ on $\mathbb{C}P^n$, then $\omega = 0$.

Consider the metric

$$ds^2 = \sum_k dz^k d\bar{z}^k$$

on the sphere, which is invariant with respect to the action of $S^1$ and therefore generates a metric on the quotient space $\mathbb{C}P^n = S^{2n+1}/S^1$ called the *Fubini–Study metric.*

Let $\Omega$ be the imaginary part of this metric on $\mathbb{C}P^n$. On the sphere $S^{2n+1}$ we have

$$\pi^*\Omega = \frac{i}{2}\Big(\sum_k dz^k \wedge d\bar{z}^k\Big).$$

Obviously, this form is closed on the sphere $S^{2n+1}$ (and even exact: $\sum_k dz^k \wedge d\bar{z}^k = d\big(\sum_k z^k d\bar{z}^k\big)$). Therefore, $d(\pi^*\Omega) = \pi^*(d\Omega) = 0$, and we conclude that $d\Omega = 0$. Thus we have proved the following theorem.

**Theorem 11.7.** *Complex projective spaces $\mathbb{C}P^n$ with Fubini–Study metric are Kählerian.*

In what follows we will assume that the Kähler form corresponding to the Fubini–Study metric is normalized so that its integral over the submanifold $\mathbb{C}P^1 = \{(z_1 : z_2 : 0 : \cdots : 0)\}$ equals one.

In 5.1.7 we defined algebraic manifolds as complex-analytic submanifolds in $\mathbb{C}P^n$, $n \geq 1$. Theorems 11.6 and 11.7 imply the following result.

**Corollary 11.4.** *Algebraic manifolds (with metrics induced by the imbedding into $\mathbb{C}P^n$) are Kählerian.*

Let $M$ be a compact Kähler manifold. As any complex manifold, it is oriented. At each point, the Kähler form $\Omega$ can be written in appropriate coordinates as $\Omega = \sum_{k=1}^n dx^k \wedge dy^k$; therefore,

$$\frac{1}{n!}\,\Omega^n = (dx^1 \wedge dy^1) \wedge \cdots \wedge (dx^n \wedge dy^n).$$

We see that the form $\frac{1}{n!}\,\Omega^n$ is the volume form on the manifold, hence

$$\int_M \Omega^n > 0.$$

Thus the form $\Omega$ is not exact and realizes a nontrivial cohomology class

$$[\Omega^n] = [\Omega]^n \neq 0;$$

otherwise, by the Stokes theorem, we would have $\int_M \Omega^n = \int_{\partial M} \alpha = 0$, where $d\alpha = \Omega^n$. We have thereby proved the following theorem.

**Theorem 11.8.** *The Kähler form $\Omega$ on a compact Kähler manifold of complex dimension $n$ is not cohomological to zero, neither are its first $n$ powers,*

$$[\Omega]^k \neq 0, \quad k = 1, \dots, n.$$

*In particular, the Betti numbers $b_{2p} = \dim H^{2p}(M; \mathbb{R})$ of even degree are positive.*

**Corollary 11.5.** *Complex-analytic submanifolds of Kähler manifolds are not homological to zero.*

**Proof.** Let $\varphi\colon N \to M$ be an imbedding, and let $\Omega$ be a Kähler form on $M$. Denote by $k$ the complex dimension of the submanifold $N$. We know that

$$\int_{\varphi(N)} \Omega^k = \int_N \varphi^* \Omega^k > 0.$$

If a cycle $(N, \varphi)$ is homological to zero, then the integrals over it of all closed forms vanish (see Theorem 9.11). Hence the corollary. □

**Corollary 11.6.** *There exist compact complex manifolds that have no Kähler structure. This is true, e.g., for $S^1 \times S^3$.*

**Proof.** On the space $\mathbb{C}^2 \setminus \{0\}$, consider the complex-analytic action of the group $\mathbb{Z}$ generated by the mapping $(z_1, z_2) \to (2z_1, 2z_2)$. The quotient space $M^4$ is a complex manifold diffeomorphic to the Cartesian product of the spheres $S^1 \times S^3$. Since these spheres are Lie groups, $S^1 = \mathrm{U}(1)$ and $S^3 = \mathrm{SU}(2)$, any cohomology class of the homogeneous space $S^1 \times S^3$ is realized by an invariant form, which is expressed in terms of invariant forms on the spheres $S^1$ and $S^3$. Therefore, any cohomology class of $M^4$ is representable as a sum of products of cohomology classes of the spheres $S^1$ and $S^3$. Since $H^1(S^3) = H^2(S^3) = H^2(S^1) = H^2(S^3) = 0$, we have $H^2(S^1 \times S^3) = 0$, hence there is no Kähler form on the manifold $S^1 \times S^3$. □

With the aid of Hodge theory for compact Kähler manifolds, one can prove a number of strong results, which we state for reference without proof.

1) $\dim_{\mathbb{C}} H^p_{\bar\partial}(M) < \infty$, $p = 1, \dots, 2n$, $\dim_{\mathbb{C}} M = n$.

2) The Dolbeault cohomology groups coincide with quotient spaces of closed forms modulo exact forms, i.e., with the de Rham cohomology for complex-valued forms:

$$H^p_{\bar\partial}(M) = H^p(M; \mathbb{C}) = H^p(M; \mathbb{R}) \otimes \mathbb{C}$$

(the last equality means that the group $H^p(M;\mathbb{C})$ is generated over the field $\mathbb{C}$ by the generators of the de Rham cohomology group $H^p(M;\mathbb{R})$ over the field $\mathbb{R}$), which implies the following decomposition of the Betti numbers $b_p = \dim H^p(M;\mathbb{R})$:

$$b_p = \sum_{r+s=p} h_{r,s}, \quad h_{r,s} = \dim_{\mathbb{C}} H^{r,s}_{\bar\partial}(M). \tag{11.13}$$

3) The complex conjugation of forms $\omega \to \bar\omega$ generates isomorphisms of cohomology groups

$$H^{r,s}_{\bar\partial}(M) = \overline{H^{s,r}_{\bar\partial}(M)},$$

which implies the equality $h_{r,s} = h_{s,r}$.

4) The Betti numbers $b_{2k+1}$ of odd degree are always even (which follows from the decomposition (11.13) and the equality $h_{r,s} = h_{s,r}$).

**11.2.4. Almost complex structures.** Let $M$ be a smooth manifold, and suppose that in the tangent space at each point of $M$ there is an isomorphism

$$J\colon T_xM \to T_xM$$

such that $J^2 = -1$. Moreover, assume that the tensor field $J = (a^k_l)$ is smooth.

In this case the manifold is said to be equipped with an *almost complex structure*, and the manifold itself is said to be *almost complex*.

**Lemma 11.4.** *Any almost complex manifold is of even dimension.*

**Proof.** Let $J\colon \mathbb{R}^p \to \mathbb{R}^p$ be a linear transformation of a finite-dimensional vector space such that $J^2 = -1$. We have $\det J^2 = (-1)^p$, and, on the other hand, $\det J^2 = (\det J)^2 \geq 0$. Therefore, $\det J^2 = 1$ and $p = 2n$. Hence the lemma. □

As an example of almost complex manifolds we may consider complex manifolds. On these manifolds, complex-valued forms of type $(1,0)$ are generated at each point, over the field of real numbers $\mathbb{R}$, by the forms

$$\alpha^k = dx^k + iJ(dx^k), \quad k = 1, \dots, 2n,$$

whereas the forms

$$\beta^k = dx^k - iJ(dx^k), \quad k = 1, \dots, 2n,$$

constitute a basis of forms of type $(0,1)$.

An almost complex structure is said to be *integrable* if there are complex coordinates on the manifold for which it is a complex structure.

**Theorem 11.9.** *An almost complex structure $J$ is integrable if and only if the tensor*

$$N^i_{kl} = a^j_k\Big(\frac{\partial a^i_l}{\partial x^j} - \frac{\partial a^i_j}{\partial x^l}\Big) - a^j_l\Big(\frac{\partial a^i_k}{\partial x^j} - \frac{\partial a^i_j}{\partial x^k}\Big), \tag{11.14}$$

*where $J = (a^i_k)$, vanishes everywhere:*

$$N^i_{kl} = 0.$$

**Proof.** We will prove only the necessity of the condition $N^i_{kl} = 0$, $i, k, l = 1, \dots, n$, since the converse statement requires a complicated analytic proof.

If the structure $J$ is integrable, then the forms $\alpha^k$, $k = 1, \dots, 2n$, are linear combinations of the forms $dz^j$, $i\,dz^j$, $j = 1, \dots, n$; hence their differentials $d\alpha^k$ can be written as

$$d\alpha^k = \sum_l \omega^l \wedge \alpha^l.$$

Note that for any vector $\eta$,

$$\alpha^k(\eta - iJ\eta) = 0, \quad k = 1, \dots, 2n.$$

Therefore, if the structure is integrable, then for any pair of vectors $\xi$ and $\eta$ the following equalities hold:

$$d\alpha^k(\xi - iJ\xi, \eta - iJ\eta) = \sum_l \omega^l \wedge \alpha^l(\xi - iJ\xi, \eta - iJ\eta) = 0, \quad k = 1, \dots, 2n.$$

We write them down in more detail:

$$d\alpha^k = d(dx^k + ia^k_l\,dx^l) = i\sum_{l<m}\Big(\frac{\partial a^k_m}{\partial x^l} - \frac{\partial a^k_l}{\partial x^m}\Big)\,dx^l \wedge dx^m,$$

$$d\alpha^k(\xi - iJ\xi, \eta - iJ\eta) = i\sum_{l<m}\Big(\frac{\partial a^k_m}{\partial x^l} - \frac{\partial a^k_l}{\partial x^m}\Big)(\delta^l_p - ia^l_p)(\delta^m_q - ia^m_q)\xi^p\eta^q = 0.$$

We leave it as a simple exercise to show by a simple calculation that these equalities are equivalent to $N^i_{kl} = 0$ for all $i, k, l$. Thus the necessity of the condition $N = 0$ for integrability of an almost complex structure is proved. □

The tensor $N$ specified by formula (11.14) is called the *Nijenhuis tensor.*

EXAMPLE. We will give an example of a nonintegrable almost complex structure on the 6-dimensional sphere $S^6$. To this end we introduce the *algebra of octonions* (= *octaves*) (or the *Cayley algebra*) $\mathbb{O}$. It is constructed from the algebra of quaternions as follows.

We set $\mathbb{O} = \mathbb{H}^2$ and define multiplication by the formula

$$x \cdot y = (x_1, x_2) \cdot (y_1, y_2) = (x_1y_1 - \overline{y_2}x_2, x_2\overline{y_1} + y_2x_1).$$

We specify conjugation on the algebra $\mathbb{O}$ by the rule $\overline{(x_1, x_2)} = (\overline{x_1}, -x_2)$.

Note that by this rule we can successively introduce, starting with $\mathbb{R}$, the algebras $\mathbb{C} = \mathbb{R}^2$ and $\mathbb{H} = \mathbb{C}^2$.

The algebra of octaves is a division algebra: to each nonzero element uniquely corresponds its inverse (we will not verify this). Multiplication in this algebra is not associative, i.e., the identity $a(bc) = (ab)c$ does not always hold. However, it holds in any subalgebra generated by any two elements of $\mathbb{O}$.

On the algebra of octaves $\mathbb{O} = \mathbb{R}^8$, we define the Euclidean scalar product as

$$(x, y) = \operatorname{Re} x\bar{y}.$$

The imaginary octaves (i.e., those satisfying the relation $\bar{x} = -x$) form a 7-dimensional subspace, in which we consider the unit sphere $S^6$. We see that in $\mathbb{O}$ this sphere is given by the equations

$$\bar{x} = -x, \quad (x, x) = -x^2 = 1, \quad x \in \mathbb{O} = \mathbb{R}^8. \tag{11.15}$$

Now we will construct an almost complex structure on this sphere. Each tangent vector to the sphere is represented by an imaginary octave $\xi$, i.e., by a vector in $\mathbb{R}^7$.

**Lemma 11.5.** *The formula*

$$J(\xi) = \xi x, \quad \xi \in T_x S^6, \tag{11.16}$$

*specifies an almost complex structure on the* 6*-dimensional sphere* (11.15).

**Proof.** Since $J\xi = \xi x = -\xi\bar{x}$ and the tangency condition is $(\xi, x) = \operatorname{Re} \xi\bar{x} = 0$, we see that $J\xi \in \operatorname{Im} \mathbb{O} = \mathbb{R}^7$. We have

$$(J\xi, x) = \operatorname{Re}(J\xi)\bar{x} = -\operatorname{Re}(J\xi)x = -\operatorname{Re}(\xi x)x = -\operatorname{Re} \xi(x^2) = \operatorname{Re} \xi = 0.$$

Therefore, $J(T_x S^6) \subset T_x S^6$, and since $J^2\xi = (\xi x)x = \xi(x^2) = -\xi$, we conclude that $J$ is an almost complex structure on $S^6$. $\square$

One can verify that the Nijenhuis tensor of this structure is not equal to zero, hence it is not integrable (see problems below).

It can be proved by topological methods that among even-dimensional spheres, only $S^2$ and $S^6$ have almost complex structures. The two-dimensional sphere is a complex manifold, $S^2 = \mathbb{C}P^1$. The question whether there exists a complex structure on $S^6$ is open so far.

Moreover, it was proved, also by topological methods, that only for dimensions 1, 2, 4, and 8 there exist (real) division algebras. These possibilities are realized by the algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$, respectively. As a consequence, only for the spheres of dimension $n = 1, 3, 7$ the tangent bundle $TS^n$ is homeomorphic to the Cartesian product $S^n \times \mathbb{R}^n$.

Note that a manifold $M^n$ is said to be *parallelizable* if its tangent bundle is trivial, i.e., homeomorphic to the Cartesian product: $TM^n = M^n \times \mathbb{R}^n$.

**11.2.5. Abelian tori.** In 5.1.7 we defined Abelian tori as complex tori $\mathbb{C}^n/\Gamma$ that can be imbedded complex-analytically into a complex projective space $\mathbb{C}P^N$ of sufficiently large dimension. Under such an imbedding, the Fubini–Study metric induces a Kähler metric on the torus.

**Lemma 11.6.** *Let $f\colon M \to \mathbb{C}P^n$ be a complex-analytic imbedding, and let $\Omega = f^*\Omega_0$ be the Kähler form on $M$ induced by the Fubini–Study form on $\mathbb{C}P^n$ under this imbedding. Then the integrals of $\Omega$ over all two-dimensional cycles in $M$ are integer-valued.*

**Proof.** Let $(z_1 : \cdots : z_{n+1})$ be homogeneous coordinates in $\mathbb{C}P^n$ (they are defined up to multiplication by nonzero constants $\lambda \in \mathbb{C}$). The subspace $\mathbb{C}P^{n-1}$ is specified in $\mathbb{C}P^n$ by the equation $z_{n+1} = 0$, and its complement is diffeomorphic to $\mathbb{C}^n$, since its points are uniquely parametrized by tuples $(z_1 : \cdots : z_n : 1)$.

If $g\colon N^2 \to \mathbb{C}P^n$, $n > 1$, is a smooth mapping, then it may be homotopically deformed into a mapping $g'\colon N^2 \to \mathbb{C}P^{n-1}$. Indeed, by the Sard theorem (see Section 12.1), the image of the manifold $N^2$ does not cover the entire complement $\mathbb{C}P^n \setminus \mathbb{C}P^{n-1} = \mathbb{C}^n$. Then we can choose a point $P$ that does not lie in the image $g(N^2)$, and deform $\mathbb{C}^n \setminus P$ along the rays passing through $P$ into "infinity", i.e., into $\mathbb{C}P^{n-1}$.

Repeating this procedure successively, we see that any two-dimensional cycle $(N^2, g)$ in $\mathbb{C}P^n$ deforms into a cycle $(N^2, g')$ whose image lies in $\mathbb{C}P^1 = \{(z_1 : z_2 : 0 : \cdots : 0)\}$. Since the integral of the Fubini–Study form over $\mathbb{C}P^1$ is equal to one, for any two-dimensional cycle $(N^2, h)$ in $M$ we deform the composition $fh$ into a mapping $g\colon N^2 \to \mathbb{C}P^1$ to obtain

$$\int_{N^2} h^*\Omega = \langle[\Omega], (N^2, h)\rangle = \langle[\Omega_0], (N^2, fh)\rangle$$
$$= \int_{N^2} h^* f^*\Omega_0 = \deg g \int_{\mathbb{C}P^1} \Omega_0 = \deg g,$$

where $\deg g$ is the degree of the mapping $g$ (see Section 12.1). Hence the lemma. □

Thus we arrive at an important definition: a compact Kähler manifold is said to be a *Hodge manifold* if the integral of the Kähler form over any two-dimensional cycle is integer-valued.

The above lemma says that all the algebraic manifolds are Hodge manifolds. The converse is also true: this is a difficult *Kodaira's theorem.*

**Theorem 11.10** (Riemann's criterion). *A complex torus $M^{2n} = \mathbb{C}^n/\Gamma$ is Abelian (i.e., it is an algebraic or a Hodge manifold) if and only if, in an appropriate basis, the lattice of periods of $\Gamma$ reduces to the form*

$$\Gamma = \Delta N_1 + BN_2, \quad N_1, N_2 \in \mathbb{Z}^n,$$

*where the matrix $\Delta$ is diagonal with positive integer diagonal elements,*

$$\Delta = \begin{pmatrix} \delta_1 & & 0 \\ & & \\ 0 & & \delta_n \end{pmatrix},$$

*and the matrix $B$ is symmetric with a positive definite imaginary part,*

$$B_{jk} = B_{kj}, \quad \operatorname{Im} B > 0.$$

**Proof.** By Lemma 11.6 we can prove the theorem for a Hodge torus.

Let $\Omega = \frac{i}{2} h_{jk}\, dz^j \wedge d\bar{z}^k$ be a Kähler form. We may assume that it has constant coefficients (otherwise we achieve this by averaging, which does not affect the cohomology class of the Kähler form; see Theorem 9.13).

The form $\Omega$ with constant coefficients specifies a skew-symmetric form on the lattice:

$$\Omega\colon \Gamma \times \Gamma \to \mathbb{Z}, \quad (v, w) \to \Omega(v, w).$$

Its value on a pair of vectors $v, w$ is, obviously, equal to the value of the integral of the form $\Omega$ over the torus generated by the vectors $v$ and $w$ (the order of the vectors determines orientation on the two-dimensional torus). Since $\Omega^n$ is the volume form on the torus, the form $\Omega$ on the lattice $\Gamma$ is nondegenerate.

Now we prove a technical lemma.

**Lemma 11.7.** *Any nondegenerate skew-symmetric form $Q$ on the integer-valued lattice $\Gamma \approx \mathbb{Z}^{2n}$ is specified in an appropriate basis by a matrix*

$$\begin{pmatrix} 0 & \Delta \\ -\Delta & 0 \end{pmatrix},$$

*where the matrix $\Delta = \operatorname{diag}(\delta_1, \dots, \delta_n)$ is diagonal with positive integer diagonal elements $\delta_k > 0$, $k = 1, \dots, n$.*

**Proof.** Choose the vectors $e_1, e_{n+1}$ of the lattice $\mathbb{Z}^{2n}$ so that $Q(e_1, e_{n+1}) = \delta_1$ is minimal among all positive values of $Q(u, v)$.

Note that the quantities $Q(e_1, v)$ and $Q(e_{n+1}, v)$ for any $v$ are divisible by $\delta_1$. Indeed, assume the contrary, i.e., assume that $Q(e_1, v) = \lambda > 0$ (making the change $v \to -v$ if necessary), $\delta_1 > 1$, and the greatest common divisor of $\lambda$ and $\delta_1$ equals one. Then, according to Euclid's algorithm, there exist integers $r$ and $s$ such that

$$1 = r\delta_1 + s\lambda = Q(e_1, re_{n+1} + sv) < \delta_1,$$

in contradiction with the choice of $\delta_1$.

Each vector $v \in \Gamma$ is uniquely representable as

$$v = \frac{Q(e_1, v)}{\delta_1} e_{n+1} - \frac{Q(e_{n+1}, v)}{\delta_1} e_1 + v',$$

where $v'$ belongs to the orthogonal complement $\Gamma'$ of $e_1$ and $e_{n+1}$. Thus we have obtained the orthogonal decomposition

$$\Gamma = \mathbb{Z}(e_1, e_{n+1}) + \Gamma'.$$

Now we restrict the form $Q$ to $\Gamma'$ and repeat this process. After several repetitions we construct the required basis. $\square$

Applying this lemma to our setting, we construct a basis $e_1, \dots, e_{2n}$ for the lattice $\Gamma$ in which the form $Q$ becomes as stated in the lemma.

Now we carry through the Gram–Schmidt orthogonalization of the vectors $e_1, \dots, e_n$ relative to the metric $(v, w) = \operatorname{Re}\langle v, w\rangle$, where $\langle v, w\rangle$ is the Kähler metric:

$$v_1 = e_1, \quad v_2 = e_2 - \frac{\operatorname{Re}\langle v_1, e_1\rangle}{\langle v_1, v_1\rangle} v_1, \quad \dots, \quad v_k = e_k - \sum_{j<k} \frac{\operatorname{Re}\langle v_j, e_k\rangle}{\langle v_j, v_j\rangle} v_j, \quad \dots .$$

Since $\Omega(v_j, v_k) = -\operatorname{Im}\langle v_j, v_k\rangle = 0$ for all $j, k = 1, \dots, n$, and $\operatorname{Re}\langle v_j, v_k\rangle = \delta_{jk}$, we see that

$$\langle v_j, v_k\rangle = \delta_{jk};$$

hence the vectors $v_1, \dots, v_n$ so constructed form a basis in $\mathbb{C}^n$ over the field $\mathbb{C}$.

We will write $e_1, \dots, e_n$ for the vectors $v_1, \dots, v_n$. Note that the Gram–Schmidt transformation preserves the form $Q$.

The vectors $e_1/\delta_1, \dots, e_n/\delta_n, e_{n+1}, \dots, e_{2n}$ constitute a real basis in $\mathbb{R}^{2n} = \mathbb{C}^n$ with coordinates $x_1, \dots, x_{2n}$, in which the form $\Omega$ becomes

$$\Omega = \sum_{k=1}^{n} dx_k \wedge dx_{n+k},$$

and the lattice of periods $\Gamma$ is written as

$$\Gamma = \Delta N_1 + B N_2, \quad N_1, N_2 \in \mathbb{Z}^n.$$

What can we say about the matrix $B$? It is involved in the formula for the change of coordinates,

$$\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} 1 & B \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{2n} \end{pmatrix}.$$

Consider the following chain of equalities:

$$\begin{aligned}\Omega &= \sum_{j,k=1,\dots,n} \frac{i}{2} h_{jk}\, dz_j \wedge dz_k \\ &= \sum_{j,k} \frac{i}{2} h_{jk} \Big( dx_j + \sum_l B_{jl}\, dx_{n+l} \Big) \wedge \Big( dx_k + \sum_m \bar{B}_{km}\, dx_m \Big) \\ &= \sum dx_k \wedge dx_{n+k}\end{aligned}$$

(all summations are over $1,\dots,n$). We examine the left- and right-hand sides of this chain.

1. We have

$$\frac{i}{2} \sum_{j,k} h_{jk}\, dx_j \wedge dx_k = 0,$$

which implies that $h_{jk} - h_{kj} = h_{jk} - \overline{h_{kj}} = 0$. Hence the matrix $h_{jk}$ is symmetric and real:

$$h_{jk} = h_{kj}, \quad h_{jk} = \overline{h_{jk}}.$$

2. The following relation holds:

$$\frac{i}{2} \sum_{j,k,l} h_{kj} (\bar{B}_{jl} - B_{jl})\, dx_k \wedge dx_{n+l} = \sum_q dx_q \wedge dx_{n+q},$$

which implies that the matrices $H$ and $\operatorname{Im} B$ are mutually inverse,

$$\sum_j h_{kj} \operatorname{Im} B_{jl} = \delta_{kl}.$$

3. We have

$$\sum_{j,k} h_{jk} B_{jl} \bar{B}_{km} = T_{lm} = T_{ml} = 0.$$

Substituting the identity $\bar{B}_{km} = B_{km} - 2i \operatorname{Im} B_{km}$ into this equality, we obtain

$$\begin{aligned}T_{lm} &= \sum_{j,k} h_{jk} B_{jl} (B_{km} - 2i \operatorname{Im} B_{km}) \\ &= \sum_{j,k} h_{jk} B_{jl} B_{km} - 2i \sum_j B_{jl} \Big( \sum_k h_{jk} \operatorname{Im} B_{km} \Big) \\ &= \sum_{j,k} h_{jk} B_{jl} B_{km} - 2i \sum_j B_{jl} \delta_{jm} = \sum_{j,k} h_{jk} B_{jl} B_{km} - 2i \sum_j B_{ml}.\end{aligned}$$

Since the last expression is symmetric in $l, m$, we conclude that the matrix $B_{kl}$ is symmetric. Its imaginary part is inverse to the matrix $h_{jk}$, which

is real and specifies a positive definite scalar product. Hence the matrix $\operatorname{Im} B$ specifies such a product as well. Thus we conclude that

$$B_{jk} = B_{kj}, \quad j, k = 1, \ldots, n, \quad \operatorname{Im} B > 0.$$

This shows that Hodge's tori have the form stated in the theorem.

To prove the converse statement, it suffices to construct the metric $(h_{jk}) = (\operatorname{Im} B)^{-1}$ corresponding to the matrix $B$ and, inverting the above arguments, to see that the Kähler metric $h_{jk}\, dz^j \wedge dz^k$ specifies the Hodge form. This completes the proof. □

As a corollary, it can easily be shown that for $n \geq 2$ almost all complex tori are non-Abelian (the property of being an Abelian torus is violated by a small perturbation of the lattice of periods).

Note that Lemma 11.7 in fact establishes a stronger property (which is easily deduced from the above proof): for any $k = 1, \ldots, n-1$, the coefficient $\delta_{k+1}$ is divisible by $\delta_k$. Furthermore, it is obvious from the construction that $\delta_1, \ldots, \delta_n$ are invariants (of the Kähler form). They are called the *polarization divisors.*

If $\delta_1 = \cdots = \delta_n = 1$, then the Abelian torus is said to be *principally polarized.* It is imbedded in $\mathbb{C}P^N$ by means of the *theta-function* constructed for this torus (see 5.1.7).

## Exercises to Chapter 11

1. Prove that all conformal transformations of a connected Riemannian manifold form a Lie group.

2. Construct a complex structure on the product of spheres of odd dimension $S^{2k+1} \times S^{2l+1}$, $k, l \geq 0$.

3. Let $M$ be a Kähler manifold of complex dimension $n$ with operator $J = (J_s^r)$ of complex structure ($J^2 = -1$), let $\Omega$ be the Kähler form, and $R_{jk}$ the Ricci tensor of the Kähler metric. Define the function $f$ by the formula

$$\Omega^n = f \cdot \frac{(-1)^{n(n-1)/2} i^n n!}{2^n}\, dz^1 \wedge \cdots \wedge dz^n \wedge d\bar{z}^1 \wedge \cdots \wedge d\bar{z}^n.$$

Prove that the tensor $\rho_{jk} = J_j^l R_{lk}$ specifies a closed $(1,1)$-form, which is equal to

$$\rho = -id'd'' \log f.$$

For $n = 1$ we obtain formula (11.11).

4. Let $M$ be an almost complex manifold, and $N$ the Nijenhuis tensor (11.14). Prove that for vector fields $\xi, \eta$ on the manifold, the vector field

$N(\xi, \eta) = N^i_{kl}\xi^k\eta^l$ is expressed in terms of various commutators as

$$N(\xi, \eta) = J([J\xi, \eta] + [\xi, J\eta]) + [\xi, \eta] - [J\xi, J\eta]$$

(this operation is called the *Nijenhuis bracket*).

5. Prove that the Nijenhuis tensor of an almost complex structure (11.16) on the 6-dimensional sphere is not equal to zero.

6. Prove that all the Lie groups are parallelizable.

7. Prove that all the one-dimensional complex tori are Abelian, while for $n \geq 2$ almost all $n$-dimensional complex tori are non-Abelian. Construct an explicit example of a non-Abelian two-dimensional torus.

Chapter 12

# Morse Theory and Hamiltonian Formalism

## 12.1. Elements of Morse theory

**12.1.1. Critical points of smooth functions.** A variational problem consists in finding the points where a given function $f\colon X \to \mathbb{R}$ on a space $X$ takes its extremal (maximal or minimal) values.

In the simplest case where $f\colon M^n \to \mathbb{R}$ is a smooth function on a smooth manifold $M^n$, this problem extends to finding critical points of the function $f$. Recall that a *critical point* of a smooth function $f$ is a point $x_0 \in M^n$ where the gradient of this function vanishes,

$$\operatorname{grad} f(x_0) = \Big(\frac{\partial f(x_0)}{\partial x^1}, \dots, \frac{\partial f(x_0)}{\partial x^n}\Big) = 0.$$

Since the gradient of a function is a covector, it vanishes simultaneously in all coordinates systems; hence the definition of a critical point is independent of the choice of coordinates. The value $c = f(x_0)$ of $f$ at a critical point is called a *critical value* of the function $f$.

**Lemma 12.1.** *Let $f\colon M^n \to \mathbb{R}$ be a smooth function on a smooth manifold $M^n$. Then the points at which $f$ attains its maximum and minimum values are its critical points.*

**Proof.** Let $x_0$ be a point at which $f$ takes its maximum value. Expand $f$ into Taylor series about this point:

$$f(x) = f(x_0) + \frac{\partial f}{\partial x^j}(x^j - x_0^j) + O(r^2),$$

where $r = \sqrt{\sum_{j=1}^{n}(x^j - x_0^j)^2}$. Suppose $\operatorname{grad} f = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}\right) \neq 0$ at $x_0$. Then we can find a point $x = (x^1, \dots, x^n)$ sufficiently close to $x_0$ and such that $f(x) > f(x_0)$. Indeed, if, for example, $\frac{\partial f}{\partial x^1} \neq 0$ at $x_0$, we may set $x = x_0 + (\varepsilon, 0, \dots, 0)$ with some positive and sufficiently small $\varepsilon$. This contradicts the assumption that $f$ attains its maximum at $x_0$. Hence a maximum point is a critical point of the function.

The proof for a minimum point is similar. □

Now we give other examples of critical points.

EXAMPLES. 1. A SADDLE. Consider a smooth function $f(x, y) = x^2 - y^2$ on the plane with coordinates $x$ and $y$. The only critical point is the point $x = y = 0$. The level curves corresponding to noncritical values are hyperbolas $x^2 - y^2 = \text{const} \neq 0$, while the critical level curve consists of two straight lines $x = \pm y$ that intersect at the critical point. The graph of this function in a neighborhood of the critical point looks like a saddle (see Figure 12.1).

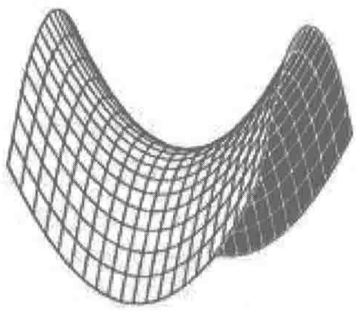

**Figure 12.1.** Saddle.

2. MONKEY SADDLE. Consider the function $f(x, y) = \operatorname{Re}(x + iy)^k$ on the plane. For $k = 2$ this is the function of the previous example. The only critical point is $x = y = 0$, and the critical level curve consists of $k$ straight lines intersecting at the critical point. These lines divide the plane into $2k$ sectors. The sectors where $f$ takes positive and negative values alternate upon traversing the plane around the point $x = y = 0$, and the level curves corresponding to noncritical values break up into $k$ curves lying in sectors with a given sign of $f$. For $k = 3$ the graph of this function has the shape of a "monkey saddle" (with three directions of descent for two legs and the tail; see Figure 12.2).

3. Consider the smooth function $f(x, y) = x^2$ on the plane with coordinates $x, y$. The critical points are not isolated and form a submanifold in $\mathbb{R}^2$, namely, the line $x = 0$. If we multiply $f(x, y)$ by $y^2$, then the set of

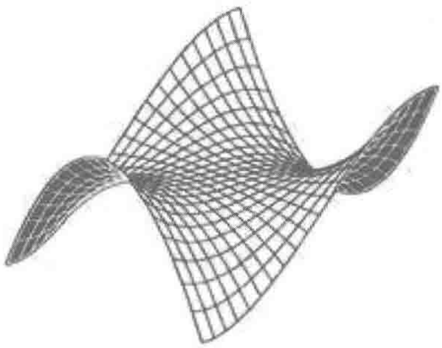

**Figure 12.2.** Monkey saddle.

critical points of the function thus obtained is not a manifold and consists of two intersecting lines $x = 0$ and $y = 0$.

4. CRITICAL POINTS OF THE RESTRICTION OF A FUNCTION TO A SUBMANIFOLD. Examples 1–3 may also be regarded as examples of critical points of the height function $h(x, y, z) = z$ restricted to submanifolds in $\mathbb{R}^3$ specified as the graphs of functions $z = f(x, y)$. In the general case the critical points of restrictions of smooth functions to manifolds are described by the following lemma.

**Lemma 12.2.** *Let $M^n$ be a smooth submanifold in a domain $U \subset \mathbb{R}^{n+k}$ specified by equations $F_1(x^1, \dots, x^{n+k}) = \dots = F_k(x^1, \dots, x^{n+k}) = 0$. Let $f\colon U \to \mathbb{R}$ be a smooth function defined in $U$, and let $g\colon M^n \to \mathbb{R}$ be the restriction of $f$ to the submanifold $M^n$. A point $x_0 \in M^n$ is a critical point of the function $g$ if and only if* $\operatorname{grad} f$ *at this point admits a linear expression in terms of* $\operatorname{grad} F_1, \dots, \operatorname{grad} F_k$.

**Proof.** This lemma follows from the Lagrange theorem on conditional extremum, familiar from a course of calculus. The Lagrange theorem says that $x_0$ is a critical point of the function $f$ restricted to $M^n$ if and only if there exist constants $\lambda_1^0, \dots, \lambda_k^0$ (*Lagrange multipliers*) such that the point $(x_0^1, \dots, x_0^{n+k}, \lambda_1^0, \dots, \lambda_k^0)$ is a critical point of the function

$$F(x^1, \dots, x^{n+k}, \lambda_1, \dots, \lambda_k) = f(x^1, \dots, x^{n+k}) + \sum_{j=1}^{k} \lambda_j F_j(x^1, \dots, x^{n+k}).$$

Indeed, in this case the point $x_0$ lies in the manifold $M^n$, since

$$\frac{\partial F}{\partial \lambda_1} = F_1 = 0, \ \dots, \ \frac{\partial F}{\partial \lambda_k} = F_k = 0,$$

and grad $f$ is expressed linearly in terms of grad $F_1, \dots,$ grad $F_k$:

$$\frac{\partial F}{\partial x^j} = \frac{\partial f}{\partial x^j} + \sum_{i=1}^{k} \lambda_i \frac{\partial F_i}{\partial x^j} = 0$$

for $j = 1, \dots, n+k$. Hence the lemma. □

As an example of application of the Lagrange theorem, we prove the following assertion.

**Theorem 12.1.** *Each symmetric matrix $A$ in $\mathbb{R}^n$ has an eigenvector with eigenvalue*

$$\lambda = \min_{|x|=1} (Ax, x).$$

**Proof.** Consider the smooth function

$$f(x) = (Ax, x)$$

and its restriction to the unit sphere $S^{n-1} = \{|x| = 1\}$.

Any smooth function on a compact manifold has a minimum point $\hat{x} \in \mathbb{R}^n$. Since the point $\hat{x}$ is critical, there is a constant $\lambda$ such that, at this point,

$$\frac{\partial}{\partial x_i} ((Ax, x) - \lambda(x, x))\Big|_{x=\hat{x}} = 0$$

for all $i = 1, \dots, n$. It follows from symmetry of $A$ that

$$(Ax, x) = \sum_{j,k=1}^{n} A_{jk} x^j x^k, \quad A_{jk} = A_{kj}, \quad \frac{\partial (Ax, x)}{\partial x_i} = 2 \sum_{k=1}^{n} A_{ik} x_k,$$

and we obtain

$$\sum_{k=1}^{n} A_{ik} x_k - \lambda x_i = 0, \quad i = 1, \dots, n,$$

i.e.,

$$A\hat{x} = \lambda \hat{x}, \quad \lambda = (A\hat{x}, \hat{x}) = \min_{|x|=1} A(x, x).$$

Hence the theorem. □

**Corollary 12.1.** *Any symmetric matrix in $\mathbb{R}^n$ is diagonalized in some orthogonal basis.*

**Proof.** We will prove this assertion by induction on the dimension of the space. For $n = 1$ it holds trivially. Assume that the corollary is proved for $n < k$.

Let $e_1$ be an eigenvector of the symmetric matrix $A$ in $\mathbb{R}^k$ (which exists by Theorem 12.1). Then the space $e_1^\perp \subset \mathbb{R}^k$ of vectors orthogonal to $e_1$ is invariant under the transformation $A$. Indeed, if $x \in e_1^\perp$, then

$$(Ax, e_1) = (x, Ae_1) = \lambda(x, e_1) = 0.$$

We have obtained an orthogonal decomposition that is invariant under the transformation $A$:

$$\mathbb{R}^k = \mathbb{R}e_1 + e_1^\perp,$$

and since $\dim e_1^\perp = k-1$, by the induction assumption the restriction of the linear transformation $A$ to $e_1^\perp$ is diagonalized in some orthogonal basis $e_2, \dots, e_k$. Then $e_1, \dots, e_k$ is the required basis in $\mathbb{R}^k$. □

The following theorem follows from the diagonalization process of a symmetric matrix.

**Theorem 12.2** (Rayleigh's principle). *Let $\lambda_1 \le \lambda_2 \le \cdots$ be the eigenvalues of a quadratic form $A$. Then the* $k$th *eigenvalue of the quadratic form equals*

$$\lambda_k = \min_{\dim L = k} \max_{x \in L,\, x \neq 0} \frac{(Ax, x)}{(x, x)},$$

*where the minimum is taken over all $k$-dimensional subspaces $L$ in $\mathbb{R}^n$.*

**12.1.2. Morse lemma and transversality theorems.** The examples show that there is a wide diversity of critical points, but only for some of them, the geometric form of the graph of the function in a neighborhood of the critical point is stable under small perturbations. For example, if we add to the function of Example 1 a sufficiently small function $g$ such that the point $x = y = 0$ is a critical point of $f + g$, then the graph of this new function will still look like a saddle, whereas if we add the function $\varepsilon y^2$ to $f(x, y) = x^2$, then the graph of the new function in a neighborhood of the critical point $x = y = 0$ will look as a saddle for arbitrarily small negative $\varepsilon$, but for positive $\varepsilon$ the point $x = y = 0$ will be a minimum point.

Thus we arrive at an important definition.

A critical point $x_0$ of a function $f\colon M^n \to \mathbb{R}$ is said to be *nondegenerate* (in the sense of Morse) if the matrix of its second-order partial derivatives is nonsingular at this point:

$$\det\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right) \neq 0.$$

This matrix is called the *Hessian.*

The Hessian of a function at a critical point is symmetric, hence it can be diagonalized. The number of negative eigenvalues of the diagonalized matrix is referred to as the *index* of the function $f$ at the critical point (the Morse index).

The following lemma shows that the notions of nondegeneracy and index of a critical point do not depend on the choice of local coordinates $(x^1, \dots, x^n)$.

**Lemma 12.3.** *At a critical point of a smooth function (and only at such points) the Hessian defines a quadratic form on the tangent space at this point.*

**Proof.** Let $(x^1, \dots, x^n)$ and $(y^1, \dots, y^n)$ be two local coordinate systems in a neighborhood of a point $x_0$. Using the chain rule, we obtain

$$\frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial}{\partial x^i}\Big(\frac{\partial f}{\partial x^j}\Big) = \frac{\partial}{\partial x^i}\Big(\frac{\partial f}{\partial y^k}\frac{\partial y^k}{\partial x^j}\Big) = \frac{\partial^2 f}{\partial y^k \partial y^l}\frac{\partial y^k}{\partial x^j}\frac{\partial y^l}{\partial x^i} + \frac{\partial f}{\partial y^k}\frac{\partial^2 y^k}{\partial x^i \partial x^j}.$$

This formula shows that the Hessian transforms as a tensor that specifies a quadratic form only if the second term in the last expression vanishes. For arbitrary local changes of coordinates, the Hessian always vanishes only at the critical points of $f$. Hence the lemma. □

This lemma clarifies the use of the term "index": in linear algebra the *index of a quadratic form* $\langle \xi, \eta \rangle$ on a vector space $V$ means the maximum dimension of subspaces $V'$ on which the quadratic form is negative definite, $\langle \xi, \xi \rangle < 0$ for $\xi \in V' \setminus \{0\}$. Obviously, this quantity equals the number of negative eigenvalues of the diagonalized matrix of the quadratic form.

We have the following important theorem.

**Theorem 12.3** (Morse). *Let $x_0$ be a nondegenerate critical point of a function $f$. Then in a neighborhood of this point, there are local coordinates $(x^1, \dots, x^n)$ such that $x_0 = (0, \dots, 0)$ and*

$$f(x) = f(x_0) - (x^1)^2 - \cdots - (x^k)^2 + (x^{k+1})^2 + \cdots + (x^n)^2$$

*(and the index of this critical point equals $k$).*

In the proof of this theorem we will use summation signs in a usual manner, as in calculus.

First we will prove the following lemma.

**Lemma 12.4.** *Let $f$ be a smooth function in a convex neighborhood $V$ of the point $0 \in \mathbb{R}^n$ and $f(0) = 0$. Then, in this neighborhood, $f$ is representable as*

$$f(x^1, \dots, x^n) = \sum_{i=1}^{n} x^i g_i(x^1, \dots, x^n),$$

*where $g_1, \dots, g_n$ are smooth functions and $g_i(0) = \frac{\partial f}{\partial x_i}(0)$, $i = 1, \dots, n$.*

**Proof.** Write $f$ as

$$f(x_1, \dots, x_n) = \int_0^1 \frac{df(tx^1, \dots, tx^n)}{dt}\, dt = \int_0^1 \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(tx^1, \dots, tx^n) x^i \, dt$$

and set

$$g_i(x^1, \dots, x^n) = \int_0^1 \frac{\partial f}{\partial x^i}(tx^1, \dots, tx^n)\, dt.$$

□

**Proof of Theorem 12.3.** We will assume for simplicity that $f(x_0) = 0$ (which is achieved by subtracting the constant $f(x_0)$ from $f(x)$) and the point $x_0$ is the origin of the coordinate system $(y^1, \dots, y^n)$. By Lemma 12.4 in some neighborhood of $x_0$ we have

$$f(y^1, \dots, y^n) = \sum_{i=1}^n y^j g_j(y^1, \dots, y^n).$$

Since $x_0$ is a critical point of $f$,

$$g_j(0) = \frac{\partial f}{\partial y^j}(0) = 0,$$

and applying Lemma 12.4 to the functions $g_1, \dots, g_n$ we obtain

$$f(y^1, \dots, y^n) = \sum_{i=1}^n y^i y^j h_{ij}(y^1, \dots, y^n),$$

where

$$g_j(y^1, \dots, y^n) = \sum_{i=1}^n y^i h_{ij}(y^1, \dots, y^n).$$

Replacing if necessary the smooth functions $h_{ij}$ by $\frac{1}{2}(h_{ij} + h_{ji})$, we can assume them to be symmetric in $i, j$.

We will construct the required coordinates by induction. Assume that we have already constructed coordinates $z^1, \dots, z^n$ in a neighborhood $U$ of $x_0$ in which

$$f = \pm(z^1)^2 \pm \cdots \pm (z^{l-1})^2 + \sum_{i,j \geq l} z^i z^j h_{ij}(z^1, \dots, z^n),$$

where the matrices $(h_{ij}(z^1, \dots, z^n))$ are symmetric. By a linear change of the last $n - l + 1$ coordinates we can achieve that $h_{ll}(0) \neq 0$. Set

$$g(z^1, \dots, z^n) = \sqrt{|h_{ll}(z^1, \dots, z^n)|}.$$

In a neighborhood $U' \subset U$ of $x_0$, where the smooth function $g$ does not vanish, we introduce the functions $u_1, \dots, u_n$ by the formulas

$$u^i = z^i, \quad i \neq l,$$

$$u^l(z^1, \dots, z^n) = g(z^1, \dots, z^n)\left[z^l + \sum_{i>l} z^i \frac{h_{il}(z^1, \dots, z^n)}{h_{ll}(z^1, \dots, z^n)}\right].$$

In view of the implicit function theorem, in a sufficiently small neighborhood $U''$ of $x_0$ these functions specify local coordinates in which

$$f = \sum_{i \le l} \pm (u^i)^2 + \sum_{i,j>l} u^i u^j h'_{ij}(u^1, \dots, u^n).$$

Applying this procedure successively, we obtain the required coordinates. The proof of Theorem 12.3 is completed. □

**Corollary 12.2.** *A nondegenerate critical point of a smooth function is isolated, i.e., in a sufficiently small neighborhood of this point there are no other critical points of this function.*

EXAMPLE 1. Let $M^{n-1}$ be a hypersurface in $\mathbb{R}^n$, and let $F\colon M^{n-1} \to \mathbb{R}$ be the coordinate function $x^n$, which is called the height function (over the hyperplane $x^n = 0$). On $M^{n-1}$ the *Gauss map* is defined, which assigns to each point of the surface the tangent plane to it:

$$G\colon M^{n-1} \to G_{n,1}.$$

Recall that the Grassmann manifold $G_{n,1}$ is diffeomorphic to the projective space $\mathbb{R}P^{n-1}$ and this diffeomorphism associates with each hyperplane the unit normal vector to it (taken up to its direction).

**Theorem 12.4.** 1. *A point $p_0 \in M^{n-1}$ is a critical point of the height function $f(p) = x^n$ if and only if the $x^n$-axis is directed along the normal to the surface at the point $p_0$.*

2. *A critical point of the height function is nondegenerate if and only if it is a regular point of the Gauss map.*

**Proof.** The first statement follows from Lemma 12.2. To prove the second, we represent the hypersurface in a neighborhood of the critical point $p_0$ as the graph of the function

$$x^n = f(x^1, \dots, x^{n-1}).$$

In this neighborhood, $x^1, \dots, x^{n-1}$ can be taken for local coordinates on the surface. Since the normal vector is directed along the $x^n$-axis, the point $p_0$ is a critical point of $f$: $\operatorname{grad} f(p_0) = 0$. A small neighborhood of the point $(0, \dots, 0, 1) \in G_{n,1}$ is identified with a neighborhood of the vector $(0, \dots, 0, 1)$ on the unit sphere $S^{n-1}$, which near this point is specified as the graph of the function

$$y^n = \sqrt{1 - (y^1)^2 - \dots - (y^{n-1})^2}.$$

We take $y^1, \dots, y^{n-1}$ for local coordinates on $G_{n,1}$.

The Gauss map in these coordinates has the form

$$(x^1, \dots, x^{n-1}) \to -\frac{1}{\sqrt{1+\left(\frac{\partial f}{\partial x^1}\right)^2+\dots+\left(\frac{\partial f}{\partial x^{n-1}}\right)^2}}\left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^{n-1}}\right).$$

Since $\operatorname{grad} f(p_0) = 0$, the Jacobi matrix $J$ of the Gauss map at this point coincides up to a sign with the Hessian of the height function $f(p) = x^n$:

$$J = -\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right).$$

Hence the theorem. □

Recalling the interpretation of the Gaussian curvature given in 3.3.1, we obtain the following result.

**Corollary 12.3.** *Let $M^2$ be a smooth surface in $\mathbb{R}^3$. The critical points of the Gauss map $M^2 \to G_{3,1} = \mathbb{R}P^2$ are precisely the points where the Gaussian curvature equals zero.*

EXAMPLE 2. Let $M^k$ be a smooth submanifold in $\mathbb{R}^n$. Consider all smooth functions $L_p\colon M^k \to \mathbb{R}$ of the form

$$L_p(x) = |x-p|^2.$$

To treat critical points of these functions, we need some notions.

The *normal bundle* $N(M^k)$ to a submanifold $M^k$ is the smooth manifold consisting of all the pairs $(x, \xi)$, where $x$ is a point of $M^k$ and $\xi$ is a vector in $\mathbb{R}^n$ orthogonal to $M^k$ at this point. This is an $n$-dimensional smooth manifold on which the smooth mapping

$$F\colon N(M^k) \to \mathbb{R}^n, \quad F(x,\xi) = x + \xi, \tag{12.1}$$

is defined. A point $q \in \mathbb{R}^n$ is called a *focal point* if $q = x + \xi$, where $(x, \xi)$ is a critical point of the mapping $F$.

**Theorem 12.5.** 1. *A point $x_0 \in M^k$ is a critical point of the function $L_p$ if and only if the vector $\xi = p - x_0$ is orthogonal to the manifold $M^k$ at the point $x_0$.*

2. *A critical point $x_0$ of the function $L_p$ is degenerate if and only if the point $p = x_0 + \xi_0$ is a focal point, the mapping $F$ of the form (12.1) has a singularity at the point $(x_0, \xi_0)$, and the index of the point $x_0$ is equal to the degeneracy degree of the Jacobi matrix of the mapping $F$ at the point $(x_0, \xi_0) \in N$.*

Recall that the degeneracy degree of a symmetric matrix is the multiplicity of the zero root of its characteristic polynomial.

**Proof.** It is a modification of the proof of the previous theorem.

1. Let$(u^1, \dots, u^k)$ be local coordinates on $M^k$. We obtain simply by differentiation that

$$\frac{\partial L_p}{\partial u^i} = 2\frac{\partial}{\partial u^i}\langle (x(u) - p, x(u) - p\rangle = 2\Big\langle \frac{\partial x(u)}{\partial u^i}, x(u) - p\Big\rangle.$$

The vectors $\frac{\partial x}{\partial u^i}$, $i = 1, \dots, k$, form a basis in the tangent space to $M^k$ at the point $x(u)$. Therefore, $x(u)$ is a critical point of the function $L_p$ if and only if the vector $\xi = x - p$ is orthogonal to all these vectors and hence is orthogonal to the submanifold $M^k$ at the point $x(u)$.

2. Let $(x_0, \xi_0)$ be a critical point of the mapping $L_p$. By differentiating $L_p$ once more we obtain the following expression for the Hessian:

$$\frac{\partial L_p}{\partial u^i \partial u^j} = 2\Big(-\Big\langle \frac{\partial^2 x}{\partial u^i \partial u^j}, \xi_0\Big\rangle + \Big\langle \frac{\partial x}{\partial u^i}, \frac{\partial x}{\partial u^j}\Big\rangle\Big).$$

Now we will find the Jacobi matrix of the mapping $F$ at the point $(x_0, \xi_0)$. Let us introduce orthonormal coordinates in $\mathbb{R}^n$ such that in a neighborhood of the critical point $x_0$ the submanifold $M^k$ is the graph of a mapping

$$x^{k+1} = \varphi_1(x^1, \dots, x^k), \dots, x^n = \varphi_{n-k}(x^1, \dots, x^k),$$

$x_0$ is a point with coordinates $(0, \dots, 0)$, the basis vectors $e_1, \dots, e_k$ are tangent to $M^k$ at $x_0$, and the vectors $e_{k+1}, \dots, e_n$ are orthogonal to $M^k$ at this point. This implies, in particular, that the point $x^1 = \dots = x^k = 0$ is critical for functions $\varphi_1, \dots, \varphi_{n-k}$.

Since the surface is specified by the equations

$$G_1(x^1, \dots, x^n) = x^{k+1} - \varphi_1(x^1, \dots, x^k) = 0,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$G_{n-k}(x^1, \dots, x^n) = x^n - \varphi_{n-k}(x^1, \dots, x^k) = 0,$$

the vectors $\operatorname{grad} G_1, \dots, \operatorname{grad} G_{n-k}$ are orthogonal to the submanifold at each of its points. Hence we can take local coordinates on the manifold $N$ to be $u^1 = x^1, \dots, u^k = x^k$, $t^1, \dots, t^{n-k}$, where $(x, \eta) \in N$, $x$ is the point of the submanifold with coordinates $(x^1, \dots, x^k)$, and $\eta = t^i \operatorname{grad} G_i$ is the vector orthogonal to $M^k$ at the point $x$.

We have

$$F(u, t) = \Big(u^1 - t^l \frac{\partial \varphi_l}{\partial u^1}, \dots, u^k - t^l \frac{\partial \varphi_l}{\partial u_k}, \varphi_1 + t^1, \dots, \varphi_{n-k} + t^{n-k}\Big),$$

and the Jacobi matrix of this mapping is

$$J = \begin{pmatrix} \partial F/\partial u \\ \partial F/\partial t \end{pmatrix} = \begin{pmatrix} \delta_{ij} - t^l \frac{\partial^2 \varphi_l}{\partial u^i \partial u^j} & \frac{\partial \varphi_i}{\partial u^j} \\ * & 1 \end{pmatrix}.$$

At the point $(x_0, \xi_0)$ we have

$$\frac{\partial \varphi_i}{\partial u^j} = 0, \quad \left\langle \frac{\partial x}{\partial u^i}, \frac{\partial x}{\partial u^j} \right\rangle = \delta_{ij}, \quad t^l \frac{\partial^2 \varphi_l}{\partial u^i \partial u^j} = \left\langle \frac{\partial^2 x}{\partial u^i \partial u^j}, \xi_0 \right\rangle,$$

and the Jacobi matrix becomes

$$J = \begin{pmatrix} \left\langle \frac{\partial x}{\partial u^i}, \frac{\partial x}{\partial u^j} \right\rangle - \left\langle \frac{\partial^2 x}{\partial u^i \partial u^j}, \xi_0 \right\rangle & 0 \\ * & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \frac{\partial^2 L_p}{\partial u^i \partial u^j} & 0 \\ * & 1 \end{pmatrix}.$$

This formula directly implies the second statement, which completes the proof. □

A smooth function $f\colon M^n \to \mathbb{R}$ for which all critical points are non-degenerate is called a *Morse function*. It turns out that on each smooth manifold there exists a Morse function, and, what is more, Morse functions are dense among all smooth functions.

In case of a compact manifold $M^n$ the fact that Morse functions are dense means that for a given finite covering $U_\alpha$ of $M^n$ and a given function $f$, for any arbitrarily small $\varepsilon > 0$ and any positive integer $N$, there exists a Morse function $g\colon M^n \to \mathbb{R}$ that approximates $f$ and its derivatives up to the $N$th order within $\varepsilon$:

$$|f - g| < \varepsilon, \quad \left| \frac{\partial^k f}{\partial x_\alpha^{i_1} \cdots \partial x_\alpha^{i_k}} - \frac{\partial^k g}{\partial x_\alpha^{i_1} \cdots \partial x_\alpha^{i_k}} \right| < \varepsilon$$

for all tuples $(i_1, \ldots, i_k)$ with $k \leq N$ and all charts $U_\alpha$.

The proof of this theorem relies on the *Sard theorem*, which is an important technical result having numerous applications. Before stating this theorem, we recall the definition of a critical point of a mapping of manifolds $F\colon M^n \to N^k$, which was given in Chapter 3.

A point $x \in M^n$ is said to be a *critical point* of a smooth mapping $F\colon M^n \to N^k$ if the image of the differential

$$F_*\colon T_x M^n \to T_{F(x)} N^k$$

does not cover the entire tangent space to $N^k$ at the point $F(x)$. In particular, if the dimension of the manifold $M^n$ is less than that of $N^k$, then all the points of such a mapping are critical. The image $F(x)$ of a critical point is called the *critical value* of the mapping $F$.

**Theorem 12.6** (Sard). *The set of critical values of an (infinitely) smooth mapping $F\colon M^n \to N^k$ has zero measure in $N^k$.*

**Corollary 12.4.** *On any smooth submanifold $M^k \subset \mathbb{R}^n$ there exists a Morse function.*

**Proof of the corollary.** By Theorem 12.5, if a point $p \in \mathbb{R}^n$ is not a focal point of the submanifold $M^k$, then the function

$$L_p(x) = |x - p|^2$$

has no degenerate critical points and hence is a Morse function on $M^k$. But the focal points are exactly the critical values of the mapping $N(M^k) \to \mathbb{R}^n$ specified by (12.1). By Theorem 12.6, such points form a set of zero measure in $\mathbb{R}^n$, hence for almost all points $p \in \mathbb{R}^n$ the function $L_p$ is a Morse function. This proves the corollary. □

Note that by Theorem 5.6 any compact closed manifold can be imbedded into Euclidean space of sufficiently large dimension, and so on this manifold there exist Morse functions. We will not prove that Morse functions are dense.

**Corollary 12.5.** *Let $F\colon M^n \to N^k$ be a smooth mapping. Then for almost all points $y$ of $N^k$ (i.e., lying in the complement of a set of zero measure), their inverse image $F^{-1}(y)$ is a smooth submanifold in $M^n$.*

**Proof.** Suppose $y_0 = (y_0^1, \dots, y_0^k) \in N^k$ is not a critical value of the mapping $F$. Then the inverse image of $y_0$ is given in a neighborhood of any of its points by the equations

$$y^1(x^1, \dots, x^n) = y_0^1, \ \dots, \ y^k(x^1, \dots, x^n) = y_0^k.$$

Since the rank of the mapping $F$ at the points of $F^{-1}(y_0)$ is equal to $k$, by the implicit function theorem these equations specify a smooth manifold of dimension $n - k$ in a neighborhood of any point of $F^{-1}(y_0)$. Hence the corollary. □

This corollary states, in particular, that by an arbitrarily small perturbation of an arbitrary point $y \in N^k$ we can obtain a point $y'$ in *general position* whose inverse image $F^{-1}(y')$ is a smooth submanifold in $M^n$.

There is another, more general setting, where analogs of these results hold.

Let $P^l \subset N^k$ be a submanifold in $N^k$. A smooth mapping $F\colon M^n \to N^k$ is said to be *transversally regular* (or *t-regular*) along $P^l$ if for any point $x$ that is mapped into $P^l$, the following equality holds:

$$T_{F(x)}N^k = f_*(T_x M^n) + T_{F(x)}P^l,$$

i.e., the tangent space at the point $F(x) \in P^l$ is generated by tangent vectors to $N^k$ and the images of tangent vectors to $M^n$ under the mapping $F_*$. When $P^l$ is a 0-dimensional submanifold (a point), this means that it is not a critical value of the mapping $F$.

The following assertion is proved similarly to Corollary 12.5.

**Lemma 12.5.** *Let $F\colon M^n \to N^k$ be a smooth mapping transversally regular along a submanifold $P^l$. Then the inverse image $F^{-1}(P^l)$ of $P^l$ is a smooth submanifold in $M^n$ of dimension $n - k + l$.*

The following assertion is also proved with the aid of the Sard theorem.

**Lemma 12.6.** *The mappings $F\colon M^n \to N^k$ that are transversally regular along a submanifold $P^l$ are dense among smooth mappings from $M^n$ into $N^k$.*

We will not give a formal definition of the notion of density here. This is a natural generalization of the corresponding notion for smooth functions, and the definition says that each mapping is specified in coordinates by smooth functions which are approximated, together with arbitrarily many derivatives, by functions that define transversally regular mappings.

We know that on each smooth manifold there is a metric $\rho$ specifying its topology. In particular, Lemma 12.6 says that for any smooth mapping $F\colon M^n \to N^k$ and any $\varepsilon > 0$ there exists a smooth mapping $G\colon M^n \to N^k$ such that

$$\rho(F(x), G(x)) < \varepsilon \quad \text{for all} \quad x \in M^n \tag{12.2}$$

and $G$ is transversally regular along $P^l$.

It turns out that two mappings sufficiently close to each other are homotopic, as the following lemma shows.

**Lemma 12.7.** *Let $N^k$ be a closed manifold. Then there exists a constant $\varepsilon_0 > 0$ such that whenever any two mappings $F, G\colon M^n \to N^k$ satisfy inequality* (12.2) *for $\varepsilon < \varepsilon_0$, they are homotopic.*

**Proof.** Choose a Riemannian metric on $N^k$. Since the manifold $N^k$ is compact, there exists a constant $\varepsilon_0$ such that any two points $x, y$ at distance less than $\varepsilon_0$ from each other can be joined by a unique shortest geodesic $\gamma_{x,y}(t)$, where $\gamma_{x,y}(0) = x$, $\gamma_{x,y}(d(x,y)) = y$, and $d(x,y)$ is the length of this geodesic. This family of shortest geodesics is also smoothly dependent on $x$ and $y$ (see Chapter 10). We will construct the homotopy between the mappings $F$ and $G$ by the following formula:

$$H(t,x) = \gamma_{F(x),G(x)}(td), \quad \text{where} \quad d = d(F(x), G(x)).$$

The mapping $H\colon [0,1] \times M^k \to N^k$ is smooth, $H(0,x) = F(x)$, and $H(1,x) = G(x)$. Hence the lemma. $\square$

A mapping $G$ that is homotopic to $F$ and close to it, is often called a small *perturbation* of the mapping $F$.

Lemma 12.6 is most often applied in the setting where the mapping $F\colon M^n \to N^k$ is an imbedding. We may also assume that $M^n$ is a submanifold in $N^k$.

Two submanifolds $M^n, P^l \subset N^k$ are said to intersect *transversally* if the imbedding of one of them, e.g., $F\colon M^n \to N^k$, is transversally regular along the other manifold $P^l$. Since a sufficiently small perturbation of an imbedding is, obviously, also an imbedding, Lemma 12.6 says that any setting may be reduced by an arbitrarily small perturbation of the imbedding $F$ to the case where the manifolds $M^n$ and $P^l$ intersect transversally.

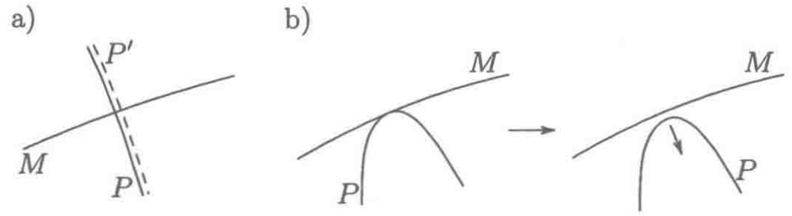


**Figure 12.3**

If $\dim N^k = \dim M^n + \dim P^l$, then the tangent spaces to $N^k$ at the points $y \in N^k$ of transversal intersection break up into the Cartesian sum of the tangent spaces to the manifolds:

$$T_yN^k = T_yM^n \oplus T_yP^l.$$

Transversal intersections (see Figure 12.3, a)) are stable relative to small perturbations in contrast to nontransversal intersections. For example, tangent submanifolds may become disjoint under a small perturbation as shown in Figure 12.3, b).

Thus, the reduction of mappings to the transversally regular form is called the reduction to the *general position*.

**12.1.3. Degree of a mapping.** Here we present briefly the notion of the degree of a mapping. Its formal justification relies on the facts given in the previous subsection 12.1.2.

Let $f\colon M \to N$ be a smooth mapping of oriented closed manifolds of the same dimension. Let $y_0$ be a regular value of this mapping. Then the *degree of the mapping* $f$ is defined by the formula

$$\deg f = \sum_{x_i \in f^{-1}(y_0)} \operatorname{sgn} \det\Big(\frac{\partial y^j}{\partial x^k}(x_i)\Big). \tag{12.3}$$

In other words, it is the sum of the signs of the Jacobian $J(f)$ taken over all points of the inverse image of $y_0$. The number of such points is finite, since they form a "0-dimensional" submanifold of the compact manifold $M$.

**Lemma 12.8.** 1. *The value of the degree of a mapping does not depend on the choice of the regular value $y_0$.*

2. *If two mappings* $f\colon M \to N$ *and* $g\colon M \to N$ *are smoothly homotopic, then these mappings have the same degrees* .

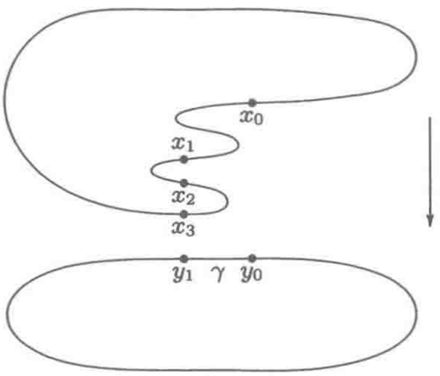


**Figure 12.4**

**Proof.** 1. Let $y_0$ and $y_1$ be two different regular values of the mapping $f$. As in Lemma 12.6, using Sard's theorem (Theorem 12.6) one can show that the points $y_0$ and $y_1$ may be joined by a path $\gamma$ along which the mapping $f$ is transversally regular (such a path may be obtained by a small perturbation from an arbitrary path $\gamma'$ joining these points). The inverse image $\Gamma = f^{-1}(\gamma)$ of this path is a smooth one-dimensional submanifold in $M$ with boundary consisting of the sets $f^{-1}(y_0)$ and $f^{-1}(y_1)$. If the points $P_0$ and $P_1$ of $f^{-1}(y_0)$ (or $f^{-1}(y_1)$) are the boundary points of the same connected component of $\Gamma' \subset \Gamma$, then continuing the function $\det J$ along the path $\Gamma'$, we see that $\det J$ has opposite signs at these points (in Figure 12.4 these are $x_2$ and $x_3$). If the points $P_0 \in f^{-1}(y_0)$ and $P_1 \in f^{-1}(y_1)$ are the boundary points of the same component of $\Gamma$, then $\det J$ has the same sign at these points (in Figure 12.4 these are $x_0$ and $x_1$). Therefore

$$\sum_{f^{-1}(y_0)} \det J = \sum_{f^{-1}(y_1)} \det J.$$

2. For the proof of homotopic invariance of the degree it suffices to repeat the same argument in a somewhat different situation. Take a smooth homotopy $F\colon M \times [0,1] \to N$, where $F(x,0) = f(x)$ and $F(x,1) = g(x)$, and choose a point $y_0 \in N$ that is a regular value of the mapping $F$. The inverse image of this point is a one-dimensional manifold $\Gamma$ in the cylinder $M \times [0,1]$ with boundary consisting of the subsets $f^{-1}(y_0)$ for $t = 0$ and $g^{-1}(y_0)$ for $t = 1$. As in the proof of part 1, we see that if two points $P_0$ and $P_1$ are the boundary points of a component of $\Gamma$ and lie on the same base of the cylinder, then they contribute opposite signs into the expressions for the degree of the mappings $f$ or $g$, whereas if they lie on the opposite bases of

the cylinder, they contribute the same sign (see Figure 12.5). This implies that $\deg f = \deg g$. Hence the lemma. □

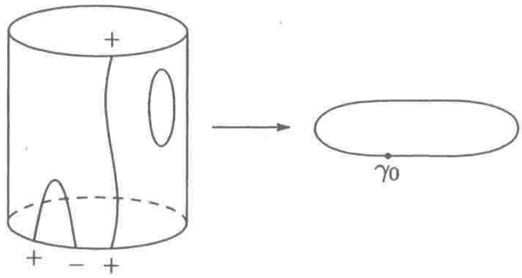


**Figure 12.5**

EXAMPLE. THE FUNDAMENTAL THEOREM OF ALGEBRA. Let $S^2$ be the two-dimensional sphere specified as the complex plane with parameter $z$ completed with the "point at infinity" (see 4.2.1). Polynomials $P(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$ determine smooth mappings $S^2 \to S^2$ taking the "point at infinity" into itself. Indeed, in a neighborhood of this point we can take $w = z^{-1}$ for a local parameter, and in terms of this parameter this mapping is written as

$$w' = \frac{1}{P(z)} = \frac{w^n}{1 + a_1 w + \cdots + a_{n-1}w^{n-1} + a_n w^n}.$$

Obviously, it is smooth in a neighborhood of the point $w = 0$.

**Lemma 12.9.** *The degree of the mapping $S^2 \to S^2$ specified by the polynomial $P(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$ is equal to $n$, i.e., to the degree of the polynomial.*

**Proof.** Let $P(z) = z^n$. Then the point $z = 1$ is regular and has $n$ inverse images. Since the mapping is complex-analytic, it preserves orientation, so that at each point of $P^{-1}(1)$ (at each $n$th root of one) the Jacobian of the mapping is positive. Hence $\deg z^n = n$.

Now it remains to note that there exists a homotopy $F(z,t) = z^n + t(a_1 z^{n-1} + \cdots + a_{n-1}z + a_n)$ between the mappings $z^n$ and $P(z)$. Hence the lemma. □

**Corollary 12.6** (Gauss theorem). *Each polynomial $P(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$ has a root in the complex plane.*

**Proof.** We will prove this by contradiction. Suppose the polynomial has no roots. Then the point $w = 0$ lies in the image of the mapping $S^2 \to S^2$ specified by this polynomial, and hence it is regular. Applying formula (12.3) to this point we obtain that $\deg P = 0$. Thus we arrive at a contradiction, which proves the corollary. □

Note that all the proofs of this theorem known so far are based on this topological argument. No purely algebraic proof of the "fundamental theorem of algebra" has been given yet.

The following theorem is useful in applications.

**Theorem 12.7.** *For a smooth mapping $f\colon M \to N$ of manifolds of the same dimension,*

$$\int_M f^*\Omega = \deg f \int_N \Omega.$$

**Proof.** The set $C$ of critical values of $f$ has zero measure in $N$, and the Jacobian is degenerate on its inverse image. Therefore

$$\int_{f^{-1}(C)} f^*\Omega = \int_C \Omega = 0.$$

Let $y_0$ be a regular value of the mapping $f$, and let $U$ be its neighborhood such that its inverse image breaks up into the union of domains $U_j$, $j = 1, \dots, m$, on which the mapping $f$ acts as a diffeomorphism on $U$: $f^{-1}(U) = U_1 \cup \dots \cup U_k$ (the existence of such a neighborhood follows from the inverse function theorem). By the theorem on the change of variable in the integral we have

$$\int_{U_i} \omega(y(x)) \det\Big(\frac{\partial y^j}{\partial x^k}\Big)\, dx^1 \wedge \dots \wedge dx^n = \operatorname{sgn} \det\Big(\frac{\partial y^j}{\partial x^k}\Big) \int_U \omega(y)\, dy^1 \wedge \dots \wedge dy^n.$$

Therefore

$$\int_{f^{-1}(U)} f^*\Omega = \deg f \int_U \Omega.$$

Since the set of regular values of the mapping $f$ is exhausted by unions of such domains $U$, the theorem follows by additivity of the integral. □

REMARK. The degree may also be defined for mappings of nonorientable manifolds. In this case the degree is the number of points in the inverse image of the regular value taken modulo 2. It is well defined, takes values in $\mathbb{Z}_2 = \{0, 1\}$, and is homotopically invariant (the above proofs are even simplified in this case).

**12.1.4. Gradient systems and Morse surgeries.** For Euclidean spaces, there is not much difference between vectors and covectors, since the index raising is done by means of the constant Euclidean metric $g_{ij} = \delta_{ij}$. In this case, by the gradient, one often means the vector pointed in the direction of the maximal growth of the function. Indeed, let $\xi$ be a tangent vector at a point $x \in \mathbb{R}^n$. The directional derivative of a function $f$ in the direction of the vector $\xi$ is given by the formula

$$\partial_\xi f = \frac{\partial f(x)}{\partial x^i} \xi^i = \langle \operatorname{grad} f(x), \xi \rangle \le |\operatorname{grad} f(x)||\xi|,$$

where $\operatorname{grad} f$ is the gradient vector field:

$$(\operatorname{grad} f)^i = \delta^{ij}\frac{\partial f}{\partial x^j}, \quad i = 1, \ldots, n.$$

For vectors of unit length (i.e., $|\xi| = 1$) the directional derivative $\partial_\xi$ in the direction of $\xi$ takes its maximum value on the vector collinear to the gradient,

$$\xi = \frac{\operatorname{grad} f}{|\operatorname{grad} f|}.$$

Accordingly, the reverse direction is the one of steepest descent of $f$.

For manifolds with Riemannian metric $g_{ij}$ one can also introduce the notion of the gradient vector field $\operatorname{grad} f$ for a function $f$ by the formula

$$(\operatorname{grad} f)^i = g^{ij}\frac{\partial f}{\partial x^j}.$$

It can be easily verified that the directional derivative of $f$ in the direction of $\xi$ is given by the formula

$$\partial_\xi f = \langle \operatorname{grad} f, \xi\rangle,$$

where the scalar product is specified by the Riemannian metric $g_{ij}$.

Suppose a manifold $M^n$ is endowed with a scalar product of covectors (which need not be positive definite and nondegenerate)

$$\langle \xi, \eta\rangle = h^{ij}\xi_i\eta_j.$$

The equation

$$\dot{x}^i = -h^{ij}\frac{\partial f}{\partial x^j}$$

is referred to as the *gradient system.* If the scalar product is specified by a smooth tensor $h^{ij}$ (which will be assumed throughout) and the manifold $M^n$ is compact, then by Theorem 10.16, for any point $x_0 \in M^n$ there exists a unique trajectory $x(t)$ of this system with initial condition $x(0) = x_0$. This solution is defined for any positive time $t \in \mathbb{R}$. Compactness of the manifold is required to guarantee that no trajectory will go "into infinity" in finite time.

The simplest example of a gradient system for a given function $f$ is specified by a Riemannian metric $g_{ij}$ and has the form

$$\dot{x}^i = -g^{ij}\frac{\partial f}{\partial x^j}.$$

As the following lemma shows, such gradient systems can be used for finding critical points of $f$.

**Lemma 12.10.** *Let $f\colon M^n \to \mathbb{R}$ be a smooth function on a closed (compact and borderless) manifold $M^n$, and let $x_0 \in M^n$ and* $\operatorname{grad} f(x_0) \neq 0$. *Then*

*any accumulation point of the trajectory $x(t)$ with initial point $x_0 = x(0)$ is a critical point of $f$.*

**Proof.** Since
$$\frac{df(x(t))}{dt} = -g^{ij}\,\frac{\partial f(x(t))}{\partial x^i}\,\frac{\partial f(x(t))}{\partial x^j} = -|\operatorname{grad} f(x(t))|^2 < 0,$$
the function $f$ is strictly decreasing along the trajectory. Compactness of the set $\{f \le f(x_0)\}$ implies that there exists an accumulation point $x_\infty$ and $f(x_\infty) < f(x(t))$ for all $t$. If $\operatorname{grad} f(x_\infty) \neq 0$, then the trajectory could be extended beyond this point, which would contradict its choice. Therefore, $\operatorname{grad} f(x_\infty) = 0$. Hence the lemma. □

For a smooth function $f\colon M^n \to \mathbb{R}$ defined on a smooth manifold $M^n$, we will denote by $M_a$ the set of points of $M^n$ where $f$ is no greater than $a$:
$$x \in M_a \subset M^n \quad \text{if} \quad f(x) \le a.$$
If $M_a$ is nonempty and $a$ is a regular value of $f$, then$M_a$ is a smooth manifold with boundary $f^{-1}(a)$.

If the function has no critical points with values in $[a, b]$, then the inverse image $f^{-1}([a, b])$ of this interval has a fairly simple structure.

**Lemma 12.11.** *Let $f\colon M^n \to \mathbb{R}$ be a smooth function on a closed manifold. Suppose that the set $f^{-1}([a, b])$ contains no critical points of $f$. Then there exists a diffeomorphism*
$$f^{-1}(b) \times [a, b] = f^{-1}([a, b])$$
*with the layers of this Cartesian product for various fixed $t \in [a, b]$ being the level surfaces:*
$$f(x, t) = t, \quad x \in f^{-1}(b),\ t \in [a, b].$$
*In particular, any level surfaces $f^{-1}(t_1)$ and $f^{-1}(t_2)$, where $t_1, t_2 \in [a, b]$, are pairwise diffeomorphic.*

**Proof.** We introduce the covector field $\xi = \operatorname{grad} f / |\operatorname{grad} f|^2$ on $f^{-1}[a, b]$ and construct the corresponding gradient system
$$\dot{x}^i = -g^{ij}\xi_j = -\frac{g^{ij}}{|\operatorname{grad} f|^2}\,\frac{\partial f}{\partial x^j}. \tag{12.4}$$
On $f^{-1}(b)$ we define the mappings
$$x \to \varphi(x, t),$$
where $\varphi(x, t)$ is the trajectory of this gradient system with initial data $\varphi(x, 0) = x$. Since
$$\frac{df(\varphi(x, t))}{dt} = -1,$$

the mapping $\varphi(x,t)$ is well defined for $0 \le t \le b-a$, and $\varphi(x,t) \in f^{-1}(b-t)$. Moreover, backward translation along the trajectories shows that any point $y \in f^{-1}(b-t)$ is representable as $\varphi(x,t)$ for a single point $x \in f^{-1}(b)$. Hence the mapping

$$f^{-1}(b) \times [a,b] \to f^{-1}([a,b]) \colon (x,t) \to \varphi(x, b-t)$$

specifies the required diffeomorphism. For $a \le c \le b$, each level surface $f^{-1}(c)$ is obtained from $f^{-1}(b)$ by translation along the trajectories of the gradient system (12.4) in time $(b-c)$ (see Figure 12.6). Hence the lemma. □

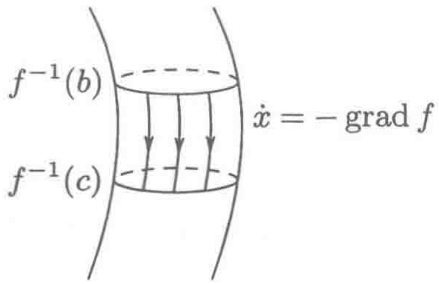


**Figure 12.6**

**Theorem 12.8.** *If on a closed manifold $M^n$ there exists a smooth function $f\colon M^n \to \mathbb{R}$ with only two critical points, then this manifold is homeomorphic to the sphere $S^n$.*

**Proof.** We will prove this theorem assuming additionally that both critical points are nondegenerate (in the general case the statement remains true, but the proof is more complicated). A continuous function on a compact manifold attains its maximum and minimum, and the maximum and minimum points are critical points. Since there are two critical points, these are precisely the maximum point $x_+$ and the minimum point $x_-$. By Theorem 12.3, in small neighborhoods of these points there exist coordinates $y_\pm^1, \dots, y_\pm^n$ in which $f$ is representable as

$$f(y) = f(x_\pm) \mp ((y_\pm^1)^2 + \cdots + (y_\pm^n)^2)$$

and the critical points have coordinates $y_\pm(x_\pm) = (0, \dots, 0)$. Hence, for sufficiently small $\varepsilon$, the submanifolds $M_{f(x_-)+\varepsilon}$ and $\{f \ge f(x_+) - \varepsilon\}$ are disjoint and homeomorphic to $n$-dimensional disks. Their boundaries are homeomorphic to the $(n-1)$-dimensional sphere $S^{n-1}$. The set $f^{-1}([f(x_-)+\varepsilon, f(x_+)-\varepsilon])$ joining them contains no critical points and is diffeomorphic to the Cartesian product $S^{n-1} \times [0,1]$. Thus the manifold $M^n$ is homeomorphic to the sphere $S^n$, which is obtained by gluing together two $n$-dimensional disks along a homeomorphism of the boundary. Hence the theorem. □

Note that this theorem does not say that the manifold $M^n$ is diffeomorphic to the standard $n$-dimensional sphere. In fact, this is not true. For example, as was shown by Milnor, on some 7-dimensional smooth manifolds there exist smooth functions with only two (nondegenerate) critical points; these manifolds are homeomorphic, but not diffeomorphic to the sphere $S^7$.

If a level surface $f^{-1}(c)$ contains a critical point of $f$, then, as a rule, close level surfaces $f^{-1}(c-\varepsilon)$ and $f^{-1}(c+\varepsilon)$ are not diffeomorphic to each other. If the critical points of $f$ are nondegenerate, then the difference between their topologies may be described explicitly.

Before doing this, we introduce two important definitions.

Let $M^n$ be a closed manifold with a $k$-dimensional sphere $S^k$ imbedded into it. Suppose this sphere has a closed *tubular neighborhood* $W$ diffeomorphic to the Cartesian product of the sphere and $(n-k)$-dimensional disks, $W = S^k \times D^{n-k}$. The boundary of this neighborhood is diffeomorphic to the Cartesian product of the spheres $S^k \times S^{n-k-1}$. In turn, this Cartesian product is also the boundary of another manifold with a boundary, namely, $W_1 = D^{k+1} \times S^{n-k-1}$. Now we delete from $M^n$ the interior of the manifold $W$ and glue the manifold $W_1$ to the boundary arisen. Thus we obtain a new closed manifold $M_1^n$. This transformation $M^n \to M_1^n$ is called *Morse surgery* or *spherical surgery* over a $k$-dimensional sphere.

EXAMPLES. 1. Let $M^2 = S^2$ be a two-dimensional sphere with imbedded 0-dimensional sphere $S^0$, which is a pair of points. Let $W$ be a tubular neighborhood of the latter sphere. Now we apply Morse surgery: we delete from the two-dimensional sphere the pair of disks, which constitute $W$, and glue the cylinder $W_1 = D^1 \times S^1$ to the boundary arisen (see Figure 12.7).

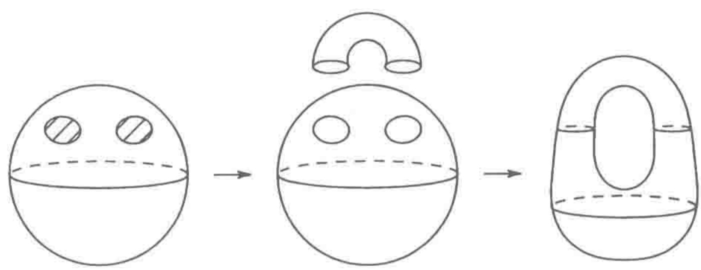

**Figure 12.7**

As a result, we obtain a two-dimensional torus $T^2$. Repeating this transformation successively, after $n$ steps we will obtain a sphere with $n$ handles. Thus we have proved the following lemma.

**Lemma 12.12.** *A sphere with $n$ handles is obtained from a two-dimensional sphere by means of $n$ Morse surgeries over 0-dimensional spheres. In particular, for $n = 1$ we obtain a two-dimensional torus $T^2$.*

2. Let $M^2 = T^2$ be a two-dimensional torus with imbedded one-dimensional sphere $S^1$, which is a closed curve such that the torus cut along this curve becomes the cylinder $Z = S^1 \times [0,1]$. We remove a tubular neighborhood of this curve and glue a tubular neighborhood $S^0 \times D^2$ of the 0-dimensional sphere instead. As a result, we obtain a two-dimensional sphere, i.e., a cylinder $Z$ each side of which is glued up with a disk (see Figure 12.8). Thus we have proved the following lemma.

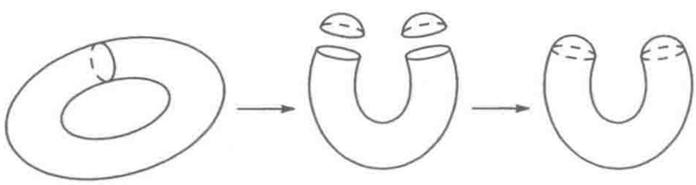

**Figure 12.8**

**Lemma 12.13.** *A two-dimensional sphere $S^2$ is obtained from the torus $T^2$ by means of Morse surgery over a one-dimensional sphere.*

Let $P^{n+1}$ be a manifold with boundary $M^n$, and let $S^{k-1}$ be a $(k-1)$-dimensional sphere imbedded into $M^n$ along with a tubular neighborhood $W = S^{k-1} \times D^{n-k+1}$. If we glue a *handle* $D^k \times D^{n-k+1}$ along this tubular neighborhood, we obtain a new manifold $P_1^{n+1}$ with a boundary. The transformation $P^{n+1} \to P_1^{n+1}$ is called the *attachment of a handle of index k*. In this case the boundary $\partial P_1^{n+1} = M_1^n$ is obtained from the boundary of the initial manifold as a result of Morse surgery over a $(k-1)$-dimensional sphere.

EXAMPLE. Let $P^3$ be a three-dimensional disk. If we attach to it a handle of index 1, we obtain the handlebody $D^2 \times S^1$.

**Theorem 12.9.** *Let $f\colon M^n \to \mathbb{R}$ be a smooth function on a closed manifold $M^n$. Suppose that the set $f^{-1}([c-\varepsilon, c+\varepsilon])$ contains a single critical point $x_0$ of $f$, $f(x_0) = c$, and this critical point is nondegenerate and has index $k$. Then:*

1) *The manifold (with a boundary) $M_{c+\varepsilon}$ is obtained from the manifold $M_{c-\varepsilon}$ by attachment of a handle of index $k$.*

2) *The level surface $f^{-1}(c+\varepsilon)$ is obtained from the level surface $f^{-1}(c-\varepsilon)$ by Morse surgery over a $(k-1)$-dimensional sphere.*

**Proof.** By Theorem 12.3 there exists a neighborhood $U$ of the point $x_0$ with coordinates $x^1, \dots, x^n$ such that $x_0 = (0, \dots, 0)$ and

$$f(x) = c - (x^1)^2 - \cdots - (x^k)^2 + (x^{k+1})^2 + \cdots + (x^n)^2$$

in this neighborhood. Take a Riemannian metric $g_{ij}$ on $M^n$ that is Euclidean in $U$, $g_{ij} = \delta_{ij}$, and take a constant $\delta > 0$ such that $\delta < \varepsilon$ and all the points with $|x| = \sqrt{(x^1)^2 + \cdots + (x^n)^2} \leq \delta$ lie in $U$. Since the Riemannian metric in $U$ is chosen to be Euclidean, the gradient field of $f$ in this domain is

$$\xi = (-2x^1, \ldots, -2x^k, 2x^{k+1}, \ldots, x^n) = (-2X, 2Y).$$

Note that by Lemma 12.11 the following diffeomorphisms hold:

$$f^{-1}(c \pm \varepsilon) \approx f^{-1}(c \pm \delta/2), \quad M_{c \pm \varepsilon} \approx M_{c \pm \delta/2}.$$

Consider the spheres $\Gamma_\pm \subset U$ specified by the equations

$$Y = (x^{k+1}, \ldots, x^n) = 0, \quad |x|^2 = \frac{\delta}{2} \quad \text{for } x \in \Gamma_-,$$
$$X = (x^1, \ldots, x^k) = 0, \quad |x|^2 = \frac{\delta}{2} \quad \text{for } x \in \Gamma_+.$$

Obviously, $\dim \Gamma_- = k - 1$, $\dim \Gamma_+ = n - k - 1$, and $\Gamma_\pm \subset f^{-1}(c \pm \delta/2)$.

The trajectories of the gradient system $\dot{x} = \operatorname{grad} f$ (see Figure 12.9) originating from the points of $\Gamma_-$ fill out the $k$-dimensional disk $D_- = \{Y = 0, |x|^2 \leq \delta/2\}$ with point $x = 0$ removed, and they converge to the critical point $x_0$ as $t \to \infty$; the trajectories that end at the points of $\Gamma_+$ fill out the $(n-k)$-dimensional disk $D_+ = \{X = 0, |x|^2 \leq \delta/2\}$ with point $x = 0$ removed, and they converge to $x_0$ as $t \to -\infty$.

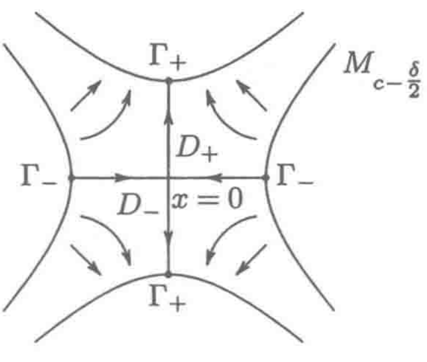


**Figure 12.9**

The complement of the disks $D_+$ and $D_-$ in $f^{-1}([c - \delta/2, c + \delta/2])$ is filled by the trajectories of the gradient system

$$\dot{x} = \frac{\operatorname{grad} f}{|\operatorname{grad} f|^2}. \tag{12.5}$$

Let $V$ be a sufficiently small tubular neighborhood of the sphere $\Gamma_-$ in the level surface $f^{-1}(c - \delta/2)$. Then translation along the trajectories of the system (12.5) establishes a diffeomorphism between the complement of $V$ in the level surface and the complement of a tubular neighborhood $\tilde{V}$ of the sphere

$\Gamma_+$ in the level surface $f^{-1}(c+\delta/2)$. Hence the level surface $f^{-1}(c+\delta/2)$ is obtained from the surface $f^{-1}(c-\delta/2)$ by removing a tubular neighborhood of a $(k-1)$-dimensional sphere and attaching a tubular neighborhood of an $(n-k-1)$-dimensional sphere instead, i.e., by means of Morse surgery over a $(k-1)$-dimensional sphere. This proves statement 2).

Let $H$ be the set of points specified by the equations

$$(x^{k+1})^2 + \cdots + (x^n)^2 \leq \frac{\delta}{4}, \quad f(x) \geq c - \frac{\delta}{2}.$$

It is homeomorphic to the $n$-dimensional disk and borders the submanifold $M_{c-\delta/2}$ along a tubular neighborhood of a $(k-1)$-dimensional sphere $\Gamma_-$ (see Figure 12.10).

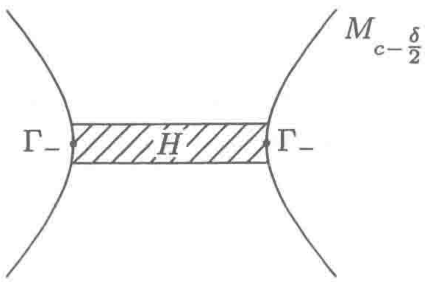


**Figure 12.10**

The set

$$S = M_{c-\delta/2} \cup H \subset M_{c+\delta/2}$$

is bounded by the surface $\partial S$, which has angles at the bordering points of $H$ with $M_{c-\delta/2}$ and hence is not smooth. It can be easily verified by a simple calculation that the gradient vector field of the function $f$ is nowhere tangent to the surface $\partial S$. On the set

$$\{f \leq c + \delta/2\} \setminus S$$

the gradient system (12.5) is well defined. With each point $y \in \partial S$ we associate the time $T(y) \leq \delta$ for which it moves along the trajectories of system (12.5) to the surface $f^{-1}(c+\delta/2)$. Now we construct the homeomorphism

$$\partial S \to f^{-1}(c+\delta/2) \colon y \to \varphi(y, T(y)),$$

where $\varphi(z,t)$ is the solution to system (12.5) with initial data $\varphi(z,0) = z$. The translations $y \to \varphi\big(y, \frac{T(y)}{\delta}\, t\big)$ establish an analog of Lemma 12.11: the closure of the set $M_{c+\delta/2} \setminus S$ is homeomorphic to the Cartesian product $\partial S \times [0, \delta]$. This implies a weakened statement 1): the manifold (with a boundary) $M_{c+\delta/2}$ is homeomorphic to the manifold $M_{c-\delta/2}$ with attached handle $H$ of index $k$.

In order to prove the diffeomorphism in statement 1), we construct a smooth surface $\partial S'$ by smoothing the angles of the boundary $\partial S$. Then the function $T(y)$ becomes smooth. It is not hard to carry this construction through, so we omit the details. □

It is seen from the proof of this theorem that all surgeries of level surfaces and submanifolds $M_a$ take place locally (in arbitrarily small neighborhoods of nondegenerate critical points). Since such points are isolated, a Morse function on a compact manifold may have only finitely many critical points, and the theorem carries over in an obvious way to the case where a critical level $f^{-1}(c)$ contains several critical points.

**Corollary 12.7.** *Let $f\colon M^n \to \mathbb{R}$ be a smooth function on a smooth manifold $M^n$. Suppose that the set $[c-\varepsilon, c+\varepsilon]$ contains a single critical value $c$. Assume that all the critical points $x_1, \dots, x_l$ lying on this critical level are nondegenerate and have indexes $k_1, \dots, k_l$. Then:*

1) *The manifold $M_{c+\varepsilon}$ is obtained from the manifold $M_{c-\varepsilon}$ by attachment of $l$ handles of indexes $k_1, \dots, k_l$.*

2) *The level surface $f^{-1}(c+\varepsilon)$ is obtained from the level surface $f^{-1}(c-\varepsilon)$ by means of Morse surgeries along $l$ spheres of dimension $k_1-1, \dots, k_l-1$, and these surgeries are carried out over disjoint domains.*

Furthermore, note that we formally described Morse surgeries and handle attachments without showing how a smooth structure is introduced on the space thus constructed. Theorem 12.9 demonstrates this by examples where surgery is related to the passage of the critical level of the Morse function (in fact, Morse surgery can always be described in this way). Although statement 2) of Theorem 12.9 follows from statement 1), we have given an independent proof, which demonstrates the connection of these surgeries with the behavior of dynamic systems in a neighborhood of a saddle critical point.

There is an extension of the Morse function, for which the surgeries of the level surfaces and submanifolds $M_a$ may be described by means of the same ideas. Namely, a function $f\colon M^n \to \mathbb{R}$ is called a *Morse–Bott function* (or simply a *Morse function*) if the following two conditions hold:

1) The critical points of the function form smooth closed submanifolds in $M^n$.

2) The critical submanifolds are nondegenerate.

Condition 2) means that for any connected critical submanifold $N^l$, there exists an integer $k \geq 0$ such that for any point $x \in N^l$ and any imbedded

disk $D$ intersecting $N^l$ transversally at the point $x$, this point $x$ is a nondegenerate critical point of index $k$ for the restriction of $f$ to the disk $D$ (this number $k$ is referred to as the *index* of the critical submanifold).

If all the critical submanifolds are 0-dimensional, i.e., each of them is a point, we obtain the definition of the Morse function.

EXAMPLE. On the Lie group $M^3 = \mathrm{SO}(3)$ of all 3-dimensional orthogonal matrices, consider the function

$$f(A) = a_1^1, \quad A = (a_j^i), \quad AA^\top = 1.$$

This function has two critical levels, $f = -1$ and $f = 1$, diffeomorphic to the circle $S^1 = \mathrm{SO}(2)$:

$$f^{-1}(-1) = \begin{pmatrix} -1 & 0 \\ 0 & -A \end{pmatrix}, \quad f^{-1}(1) = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}, \quad A \in \mathrm{SO}(2).$$

The function $f$ has no other critical values, for small positive values $\varepsilon$ the submanifolds $M_{1-\varepsilon}$ and $M_{-1+\varepsilon}$ are diffeomorphic to the handlebodies $S^1 \times D^2$, and by Lemma 12.11 the manifold SO(3) is homeomorphic to two handlebodies glued together along their boundaries.

In a similar way, the Morse–Bott functions can be constructed for matrix representations of other classical compact Lie groups. These examples are due to L. S. Pontryagin who was the first to use the Morse theory for the study of topology of particular smooth manifolds.

**12.1.5. Topology of two-dimensional manifolds.** As an example of application of Morse theory in topology, we will describe all closed two-dimensional manifolds up to homeomorphism.

Attachment of a handle to a two-dimensional manifold with boundary consists in the following.

1) If we attach a handle of index two, this means that we glue a two-dimensional disk to the manifold $M^2$ over the boundary component that is a hole homeomorphic to a circle (roughly speaking, we glue up this hole).

2) If we attach a handle of index one, this means, up to homeomorphism, that we take the square $[0,1] \times [0,1]$ and glue it along the opposite sides $0 \times [0,1]$ and $1 \times [0,1]$ to various parts of the boundary $\partial M^2$.

There are two different ways to attach a handle of index one even to a two-dimensional disk. The disk is homeomorphic to the square $D$ specified on the Euclidean plane by the inequalities $0 \leq x, y \leq 1$. Attachment of a handle of index one consists in gluing another such square to it. Up to

homeomorphism, there are two possibilities:

1) We glue together the squares $D$ and $D'$ identifying points on the parts of their boundaries by the rule

$$(0,y) \sim (0,y'), \quad (1,y) \sim (1,y'),$$

which yields the cylinder $S^1 \times [0,1]$.

2) We glue together the squares $D$ and $D'$ by the rule

$$(0,y) \sim (0,y'), \quad (1,y) \sim (1,1-y').$$

This results in a two-dimensional manifold with boundary, which is called a *Möbius strip*. It also arises if we identify the opposite sides of the square by the rule

$$(0,y) \sim (1,1-y). \tag{12.6}$$

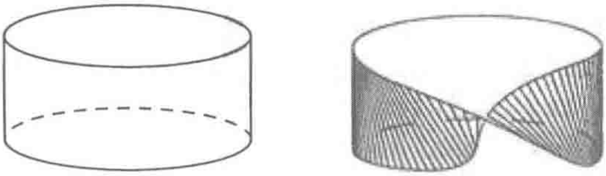

**Figure 12.11.** Cylinder and a Möbius strip.

Obviously, the cylinder and the Möbius strip are not homeomorphic, since the boundary of the former consists of two circles, whereas the boundary of the latter is a single circle.

There is an even more important difference, stated in the following lemma.

**Lemma 12.14.** *The Möbius strip is not orientable .*

**Proof.** Suppose it is orientable. Without loss of generality we may assume that the basis $e_1 = (1,0)$, $e_2 = (0,1)$ in the tangent space at the point $x = 0, y = 1/2$ is positively oriented. By continuous variation of the point $(x,y)$ we obtain positively oriented bases at various points of the square. In particular, the basis at the point $x = 1, y = 1/2$ must be positively oriented as well. But as a result of gluing (12.6) of the points $(0,1/2)$ and $(1,1/2)$ the basis $(e_1,e_2)$ at the point $(0,1/2)$ is identified with the basis $(e_1,-e_2)$ at the point $(1,1/2)$. Thus we obtain that at the point $(0,1/2)$ of the tangent space both bases $(e_1,e_2)$ and $(e_1,-e_2)$ must be positively oriented, whereas they have in fact opposite orientation. Hence the lemma. $\square$

It is easily seen that the cylinder is orientable.

In general, the following simple observation holds.

*If a two-dimensional manifold is not orientable, then imbedded in the manifold is a smooth path along which the orientation reverses, and a neighborhood of this path is homeomorphic to the Möbius strip.*

In 5.1.6 we have shown that the two-dimensional projective space $\mathbb{R}P^2$ is homeomorphic to the quotient space of the unit sphere $S^2 \subset \mathbb{R}^3$ modulo the reflection $x \to -x$. Since in this case the upper hemisphere $x^3 > 0$ is identified with the lower hemisphere, $\mathbb{R}P^2$ is homeomorphic to the space obtainable from the closed hemisphere $\{|x| = 1, x^3 \leq 0\}$ by gluing the boundary points according to the rule $x \sim -x$.

**Lemma 12.15.** *The projective plane $\mathbb{R}P^2$ is obtained from a Möbius strip by attachment of a handle of index* 2.

**Proof.** Let $I_1$ and $I_2$ be two intervals on the boundary of the hemisphere specified in polar coordinates by the inequalities $-\pi/4 < \varphi < \pi/4$ and $3\pi/4 < \varphi < 5\pi/4$. Upon identifying the boundary points by the rule $x \to -x$, these intervals will be glued into one, and one can find their neighborhoods $U_1$ and $U_2$ in the lower hemisphere which will glue together into an open two-dimensional disk $D$. The complement of $D$ is homeomorphic to the closed lower hemisphere with opposite points of two centrally symmetric intervals of the boundary glued together. But this is a Möbius strip. Therefore, having attached to its boundary the disk $D$ (a disk of index 2), we obtain the real projective plane. Hence the lemma. □

The projective plane admits the following elegant description in terms of Morse functions.

**Theorem 12.10.** *Let $M^2$ be a closed two-dimensional manifold on which there exists a Morse function with three critical points. Then this manifold is homeomorphic to $\mathbb{R}P^2$.*

**Proof.** Let $F\colon M^n \to \mathbb{R}$ be such a Morse function, and let $x_+$ and $x_-$ be its maximum and minimum points. They have indices 2 and 0, respectively. The third critical point must have index 1. We prove this by contradiction.

Suppose $x_0$ has index 0. Then it is a local minimum point, and by Lemma 12.11, for $\varepsilon > 0$ small enough, the manifold $M_{f(x_0)+\varepsilon}$ is a pair of disks, each containing a single critical point of index 0, while the manifold $\{f \geq f(x_+) - \varepsilon\}$ is homeomorphic to the disk. Since there are no other critical points, their boundaries $f^{-1}(f(x_0) + \varepsilon)$ and $f^{-1}(f(x_+) - \varepsilon)$ must be diffeomorphic by Lemma 12.11, but they consist of different numbers of circles, and so we arrive at a contradiction. The same argument with $f$ replaced by $-f$ shows that the index of $x_0$ is not equal to 2.

Therefore, $x_0$ is a critical point of index 1 and $f(x_-) < f(x_0) < f(x_+)$.

Theorem 12.9 implies that $M^2$ is obtained from the two-dimensional disk $M_{f(x_-)+\varepsilon}$ by successive attachment of handles of indexes 1 and 2. Since the manifold $M^2$ has no boundary, having attached a handle of index 1 to the disk we must obtain a manifold $N^2$ with connected boundary, which will be glued up by a disk (a handle of index 2). We have already considered two possible ways of attaching handles of index 1 to the two-dimensional disk, and we see that $N^2$ must be homeomorphic to the Möbius strip rather than to the cylinder. It remains to use Lemma 12.15 to show that $M^2$ is homeomorphic to $\mathbb{R}P^2$. Hence the theorem. □

Note that, unlike Theorem 12.8, the requirement in the above theorem that all the critical points be nondegenerate is essential: on the two-dimensional torus $T^2$ there exists a smooth function with three critical points that is not a Morse function.

Recall that the connected sum of $g$ copies of two-dimensional tori is called a *sphere with $g$ handles*. If we delete from such a sphere the interiors of $k$ pairwise disjoint imbedded disks, we obtain a sphere with $g$ handles and $k$ holes. A disk is a sphere with one hole.

The following lemma is now intuitively obvious.

**Lemma 12.16.** *Let $M$ be an orientable two-dimensional manifold which is the union of finitely many spheres with holes and handles, $M = \bigcup M_i$, and let $M'$ be the oriented manifold obtained from $M$ by attaching a handle of index* 1. *Then:*

1) *The number of connected components of the manifold $M'$ is no greater than that of $M$.*

2) *If the resulting number of connected components decreases, then $M'$ is obtained from $M$ by gluing together a sphere $S_1$ with $g_1$ handles and $k_1$ holes and a sphere $S_2$ with $g_2$ handles and $k_2$ holes into a sphere $S$ with $g_1 + g_2$ handles and $k_1 + k_2 - 1$ holes.*

3) *If the resulting number of connected components remains unchanged and the handle is attached to a single component of the boundary of $M$, then $M'$ is obtained from $M$ by a transformation of the sphere with $g$ handles and $k$ holes into a sphere with $g$ handles and $k + 1$ holes.*

4) *If the resulting number of connected components remains unchanged and the handle is attached to different components of the boundary of $M$, then $M'$ is obtained from $M$ by a transformation of the sphere with $g$ handles and $k$ holes into a sphere with $g$ handles and $k - 1$ holes.*

**Corollary 12.8.** *Attachment of a handle of index* 1 *or* 2 *to a union of spheres with handles and holes yields again a union of spheres with handles and holes.*

Note that transformations 3) and 4) do not involve imbedding Möbius strips, for which reason the numbers of handles and holes stated in the lemma are obtained.

The complete classification of oriented closed two-dimensional manifolds is given now by the following theorem.

**Theorem 12.11.** *Each orientable connected closed two-dimensional manifold is homeomorphic to a sphere with handles.*

**Proof.** Let $M^2$ be such a manifold. Take a Morse function $f$ on it and consider surgeries of the sets $M_a$ when $a$ varies from the minimum value of $f$ to its maximum.

Since nondegenerate critical points are isolated and the manifold $M^2$ is compact, the function $f$ has only finitely many critical points.

If $c_-$ is the minimum value of $f$ and the interval $[c_-, c_- + \varepsilon]$ contains no other critical points, then the submanifold $M_{c_-+\varepsilon}$ is diffeomorphic to the union of finitely many disks corresponding to the critical points for which $f = c_-$. When $a$ crosses other critical levels, we see that either two-dimensional disks corresponding to local minima of $f$ are attached to $M_a$ or handles of indexes 1 or 2 are glued. By Corollary 12.8, in either case we obtain a union of spheres with handles and holes. Since the manifold $M^2$ is connected and has no boundary, we conclude that it is homeomorphic to a sphere with handles. Hence the theorem. □

In order to complete this classification, we have to prove that spheres with different number of handles are not homeomorphic. This can be shown as follows. Let $M^2$ be a sphere with $g$ handles, and let $\gamma$ be a contour on it which has no self-intersections and does not encircle a domain on the surface. Then, on cutting the surface along this contour and gluing up the boundary with a pair of disks, we obtain a sphere with $(g-1)$ handles. By such operations we can pass from any sphere with $g$ handles to a sphere with $k$ handles, where $k < g$, but not conversely. Therefore, these surfaces are not homeomorphic.

Moreover, it is known that in the two-dimensional case homeomorphic smooth manifolds are always diffeomorphic.

The classification of nonorientable closed two-dimensional manifolds is a little more difficult to obtain, although the reasoning is similar. We only formulate the main result.

**Theorem 12.12.** *Any nonorientable connected closed two-dimensional manifold is homeomorphic to a connected sum of finitely many projective planes:*

$$M^2 = \mathbb{R}P^2 \# \cdots \# \mathbb{R}P^2.$$

*Connected sums of different numbers of projective planes are not homeomorphic.*

## 12.2. One-dimensional problems: Principle of least action

**12.2.1. Examples of functionals (geometry and mechanics). Variational derivative.** Already in the 17th century, before the laws of mechanics were formulated, arose the idea that acoustic and light signals propagate in media with variable parameters in such a way that the signal reaches its destination in minimal possible time. This fundamental principle, put forward by Fermat, which explained the law of refraction at interface, was not adopted by Newton.

Nevertheless, in the long run this approach proved to be correct. Following the ideas of Maupertuis and Euler, Lagrange finally formulated the variational "principle of least action". This approach became known as the "Lagrangian formalism". It involves a number of important constraints on the forces which can in principle act on physical systems provided their description is sufficiently complete. The further development of this approach gave rise to new kinds of geometry (symplectic and Poisson). The modern conceptions of natural laws are impossible without this geometry: for example, the laws of quantum mechanics refining upon classical mechanics cannot in principle (so far) be formulated without recourse to it.

Consider the following example. According to our conceptions, geodesics are trajectories of free motion of a particle over a surface under the constraint that the particle must stay on the surface. In Section 10.3 we defined geodesic lines $x^i = x^i(t)$ by the equation

$$\nabla_{\dot{x}}\dot{x} = 0,$$

where $\dot{x} = \left(\frac{dx^1}{dt}, \dots, \frac{dx^n}{dt}\right)$ is the velocity vector of the curve. This equation can be rewritten as

$$\ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k = 0, \quad i = 1, \dots, n.$$

For a symmetric connection $\Gamma^i_{jk}$ compatible with the Riemannian metric $g_{ij}$, geodesics are locally the shortest paths: the length of a geodesic joining two sufficiently close points is minimal among all curves joining them (see 10.3.2). Therefore, geodesics are solutions to a variational problem. Consider this question from a general point of view.

Let $L(x, \xi, t)$ be a function of a point $x = (x^1, \dots, x^n)$ and the tangent vector $\xi = (\xi^1, \dots, \xi^n)$ at this point. For a fixed pair of points $P$ and $Q$, consider all possible smooth curves $\gamma$: $x = x(t)$, $a \leq t \leq b$, joining these

points, $x(a) = P$, $x(b) = Q$. The quantity

$$S[\gamma] = \int_\gamma L(x(t), \dot{x}(t), t)\, dt$$

specifies a *functional* on the space of all such curves. It is referred to as the *action.*

Here and in what follows we denote by a dot differentiation with respect to time $t$.

For which curve $\gamma$ does the action $S[\gamma]$ take the minimum value?

Consider examples of such problems.

EXAMPLES. 1. Let the metric be Euclidean, and let

$$L(x, \dot{x}, t) = \frac{m}{2}\, \delta_{ij}\dot{x}^i\dot{x}^j - U(x)$$

be the difference between the *kinetic energy* $\frac{m}{2}\,|\dot{x}|^2$ and the *potential energy* $U(x)$ (the function $U(x)$ is called the potential). The parametrized curves minimizing the action $S$ are the trajectories of the motion of a Newtonian particle in the potential field $f_i = -\frac{\partial U}{\partial x^i}$, where the parameter $t$ on the curve is the physical time.

2. Let $g_{ij}(x)\, dx^i\, dx^j$ be a Riemannian metric. The function

$$L(x, \xi) = \frac{1}{2}\, g_{ij}(x)\xi^i\xi^j = \frac{1}{2}|\xi|^2$$

generalizes the kinetic energy of the motion of a particle in an arbitrary Riemannian manifold. The value of the action depends on the choice of the parameter on the curve.

3. For a Riemannian metric $g_{ij}(x)\, dx^i\, dx^j$, set

$$L(x, \xi) = \sqrt{g_{ij}(x)\xi^i\xi^j} = |\xi|,$$

which is the length of the tangent vector $\xi$. Then the action $S[\gamma] = \int L\, dt$ is the length of the curve $\gamma$, and its value does not depend on the choice of the parameter on the curve.

4. A light particle (photon) moves in a medium with velocity $c(x)$ at a point $x \in \mathbb{R}^n$. Set

$$L(x, \xi) = \frac{dl}{c(x)} = dt,$$

where $t$ is the actual time. The extremal curves for the functional $S$ are described by *Fermat's principle*, which states that

*Light follows the path of least time.*

At the same time the geometric meaning of the functional $S = \int L\,dt$ is the length in the new Riemannian metric $dl_F = dl/c(x)$, which has in general a nonzero curvature. From the point of view of geometry, this example is a particular case of Example 3.

We have the following theorem.

**Theorem 12.13.** *If on a curve $\gamma$: $x^i = x^i(t)$ the value of the action $S[\gamma] = \int_\gamma L(x, \dot{x}, t)\,dt$ is critical (e.g., it is minimal among all smooth curves from $P$ to $Q$), then the Euler–Lagrange equations*

$$\frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}^i}\Big) - \frac{\partial L}{\partial x^i} = 0, \quad i = 1, \dots, n, \tag{12.7}$$

*where*

$$\frac{\partial L}{\partial \dot{x}^i} = \frac{\partial L(x, \xi, t)}{\partial \xi^i}\Big|_{\xi=\dot{x}}, \quad \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}^i}\Big) = \Big(\frac{\partial^2 L}{\partial \xi^i \partial \xi^j}\ddot{x}^j + \frac{\partial^2 L}{\partial \xi^i \partial x^j}x^j + \frac{\partial^2 L}{\partial \xi^i \partial t}\Big)\Big|_{\xi=\dot{x}},$$

*are fulfilled along this curve (the derivatives of $L$ are taken with respect to the independent variables $x, \xi$, and $t$, and then their values along the curve $\gamma$ are plugged in).*

**Proof.** Let $\eta^i = \eta^i(t)$, $a \le t \le b$, be any smooth function such that $\eta^i(a) = 0$ and $\eta^i(b) = 0$. For the proof, it will suffice to consider only strongly local variations $\eta$, i.e., infinitely differentiable and vanishing identically outside a small neighborhood of the time instant $t_0$.

Consider the curves $\gamma + \varepsilon\eta$ of the form $x^i(t) + \varepsilon\eta^i(t)$, which also go from $P$ to $Q$. They are close to $\gamma$ for small $\varepsilon$. We restrict the action functional $S$ to this family and find its derivative at $\varepsilon = 0$ (at the "point" $\gamma$):

$$\lim_{\varepsilon \to 0} \frac{S[\gamma + \varepsilon\eta] - S[\gamma]}{\varepsilon} = \frac{d}{d\varepsilon}S[\gamma + \varepsilon\eta]\Big|_{\varepsilon=0}.$$

By definition, the curve $\gamma$ is a *critical point of the functional* $S[\gamma]$ (e.g., a minimum) if for any smooth strongly local vector-function $\eta(t)$,

$$\lim_{\varepsilon \to 0} \frac{S[\gamma + \varepsilon\eta] - S[\gamma]}{\varepsilon} = \frac{d}{d\varepsilon}S[\gamma + \varepsilon\eta]\Big|_{\varepsilon=0} \equiv 0.$$

We have

$$\frac{d}{d\varepsilon}S[\gamma + \varepsilon\eta]\Big|_{\varepsilon=0} = \int_a^b \Big\{\frac{\partial L}{\partial x^i}\eta^i(t) + \frac{\partial L}{\partial \xi^i}\dot{\eta}^i(t)\Big\}\,dt = 0, \tag{12.8}$$

where the integral is taken along the curve $\gamma$: $x^i = x^i(t)$, $\xi^i = \dot{x}^i(t)$. Integrating by parts we obtain

$$\int_a^b \frac{\partial L}{\partial \xi^i}\dot{\eta}^i\,dt = \frac{\partial L}{\partial \xi^i}\eta^i\Big|_{t=b} - \frac{\partial L}{\partial \xi^i}\eta^i\Big|_{t=a} - \int_a^b \eta^i \frac{d}{dt}\Big(\frac{\partial L}{\partial \xi^i}\Big)\,dt$$
$$= -\int_a^b \frac{d}{dt}\Big(\frac{\partial L}{\partial \xi^i}\Big)\eta^i\,dt,$$

since $\eta^i(a) = \eta^i(b) = 0$. We substitute this expression into (12.8) to see that if $\gamma$ is a critical point of the functional, then for any strongly local vector-function $\eta^i(t)$,

$$\frac{d}{d\varepsilon} S[\gamma + \varepsilon\eta]\Big|_{\varepsilon=0} = \int_a^b \Big[\frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}\Big]\eta^i \, dt = 0. \tag{12.9}$$

This implies that

$$\psi^i(t) = \frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} = 0, \quad i = 1, \ldots, n, \quad a < t < b.$$

Indeed, if $\psi^i(t) \neq 0$ for some values $i$ and $t = t_0$, where $a < t_0 < b$, then we take a strongly local function $\eta^i(t)$ such that the integral (12.9) does not vanish (e.g., $\eta^i(t) = \psi^i f(t)$, where $f(t)$ is a nonnegative function vanishing outside a small neighborhood of $t_0$ and such that $f(t_0) > 0$). This contradicts condition (12.9). Thus the proof is completed. □

Now we introduce the basic terminology.

1. The integrand

$$L = L(x, \xi, t) = L(x, \dot{x}, t)$$

is referred to as the *Lagrange function* or *Lagrangian*. This is a scalar function which remains unchanged under changes of coordinates.

2. Equations (12.7) as in Theorem 12.13,

$$\frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}^i}\Big) = \frac{\partial L}{\partial x^i}, \quad i = 1, \ldots, n,$$

are called the *Euler–Lagrange equations*.

3. The trajectories satisfying the Euler–Lagrange equations are called the *extremals of the functional S*.

4. The (scalar) quantity

$$E = E(x, \dot{x}, t) = E(x, \xi, t) = \xi^i \frac{\partial L}{\partial \xi^i} - L = \dot{x}^i \frac{\partial L}{\partial \dot{x}^i} - L \tag{12.10}$$

is called the *energy*.

5. The *momentum* is a covector in the $x$-space defined as

$$p_i = \frac{\partial L}{\partial \dot{x}^i} = \frac{\partial L}{\partial \xi^i}.$$

6. The covector

$$f_i = \frac{\partial L}{\partial x^i}$$

is called the *force*. In terms of the momentum and force the Euler–Lagrange equations become

$$\dot{p}_i = f_i.$$

7. The expression

$$\frac{\delta S}{\delta x^i(t)} = \frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}$$

is called the *variational derivative* of the functional $S[\gamma]$. If the functional is a function on an infinite-dimensional manifold, then the variational derivative is a functional analog of the usual partial derivative in the "direction" of $x^i(t)$ (for fixed $i$ and $t$) in the space of smooth curves.

Here we make two remarks.

1. For a given system, the Lagrangian is not uniquely defined; rather, it is defined up to addition of a total derivative. In this case the energy and momentum change along with the Lagrangian:

$$L' = L + \frac{df(x,t)}{dt}, \quad E' = E - \frac{\partial f}{\partial t}, \quad p' = p + \frac{\partial f}{\partial x}.$$

2. If the Lagrangian $L(x, \dot{x})$ is a homogeneous function of velocities $\xi = \dot{x}$ of degree one, i.e., $L(x, \lambda\xi) = \lambda L(x, \xi)$, $\lambda > 0$ (e.g., it is the length of a vector), then the energy $E$ vanishes identically and the parameter on the extremal may be chosen arbitrarily.

**12.2.2. Equations of motion (examples).** Consider the Euler–Lagrange equations (equations of motion) for the functionals in the examples given in 12.2.1.

1. For the Lagrangian

$$L(x, \dot{x}, t) = \frac{m}{2}\,\delta_{ij}\dot{x}^i\dot{x}^j - U(x),$$

the Euler–Lagrange equations are Newton's equations of the motion of a point body of mass $m$ in the potential field $U(x)$:

$$m\ddot{x}^i = -\frac{\partial U}{\partial x^i}.$$

The energy of this system is the sum of the kinetic and potential energies,

$$E = \frac{m}{2}\,|\dot{x}|^2 + U(x).$$

2. For the Lagrangian $L(x, \dot{x}) = \frac{1}{2}\,g_{ij}\dot{x}^i\dot{x}^j$, the equations of motion have the form

$$\dot{p}_k = \frac{1}{2}\frac{\partial g_{ij}}{\partial x^k}\,\dot{x}^i\dot{x}^j, \quad \text{where} \quad p_k = g_{kj}\dot{x}^j. \tag{12.11}$$

We write them down in more detail:

$$\dot{p}_k = g_{kj}\ddot{x}^j + \frac{\partial g_{kj}}{\partial x^i}\,\dot{x}^i\dot{x}^j.$$

Since $g^{mk}g_{kj} = \delta_j^m$, we have

$$\ddot{x}^m + g^{km}\Big(\frac{\partial g_{kj}}{\partial x^i} - \frac{1}{2}\,\frac{\partial g_{ij}}{\partial x^k}\Big)\dot{x}^i\dot{x}^j = 0.$$

Here we sum over $i$ and $j$, hence

$$g^{km}\,\frac{\partial g_{kj}}{\partial x^i}\,\dot{x}^i\dot{x}^j = \frac{1}{2}\,\dot{x}^i\dot{x}^j g^{km}\Big(\frac{\partial g_{ki}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i}\Big).$$

On substituting this identity into the previous equation we obtain

$$\ddot{x}^m + \Gamma_{ij}^m\dot{x}^i\dot{x}^j = 0,$$

where $\Gamma_{ij}^m = \frac{1}{2}\,g^{km}\Big(\frac{\partial g_{ki}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\Big)$ is a symmetric connection compatible with the metric $g_{ij}$.

Hence we have proved the following theorem.

**Theorem 12.14.** *The Euler–Lagrange equation for the Lagrangian*

$$L = \frac{1}{2}\,g_{ij}\dot{x}^i\dot{x}^j = \frac{1}{2}\,\langle\dot{x},\dot{x}\rangle = \frac{1}{2}\,|\dot{x}|^2$$

*coincides with the equation for geodesics of the metric $g_{ij}\,dx^i\,dx^j$.*

3. Consider the length-of-curve functional with Lagrangian $L(x,\dot{x}) = \sqrt{g_{ij}\dot{x}^i\dot{x}^j} = |\dot{x}|$. The length of the curve $S[\gamma] = \int_\gamma L\,dt$ does not depend on the parametrization of the curve. The Euler–Lagrange equations are

$$\frac{d}{dt}\bigg(\frac{g_{kj}\dot{x}^j}{\sqrt{g_{ij}\dot{x}^i\dot{x}^j}}\bigg) = \frac{\frac{\partial g_{ij}}{\partial x^k}\,\dot{x}^i\dot{x}^j}{2\sqrt{g_{ij}\dot{x}_i\dot{x}^j}},$$

and if we take the natural parameter on the curve, for which $\sqrt{g_{ij}\dot{x}^i\dot{x}^j} =$ const, they become

$$\frac{d}{dt}\,(g_{kj}\dot{x}^j) = \frac{1}{2}\,\frac{\partial g_{ij}}{\partial x^k}\,\dot{x}^i\dot{x}^j.$$

This is the equation of geodesics (12.11) obtained for naturally parametrized curves.

Thus we have proved the following theorem.

**Theorem 12.15.** *The Euler–Lagrange equations for the length-of-curve functional (i.e., $L(x,\xi) = \sqrt{g_{ij}\xi^i\xi^j}$) coincide with the equation of geodesics if the curve is parametrized by the natural parameter.*

**Corollary 12.9.** *A smooth curve that is the shortest path joining two points $P$ and $Q$ satisfies the equation of geodesics relative to the natural parameter.*

Now we indicate two extensions of the functionals considered above.

A *Finsler metric* is a function $L(x,\xi)$ specifying, at each point $x$, a norm in the tangent space at this point. This means that $L(x,\xi)$ is homogeneous of first degree as a function of the tangent vector $\xi$,

$$L(x,\lambda\xi) = \lambda L(x,\xi) \quad \text{for } \lambda > 0, \tag{12.12}$$

nonnegative, $L(x,\xi) \geq 0$, equals zero only for $\xi = 0$, and satisfies the triangle inequality

$$L(x,\xi_1+\xi_2) \leq L(x,\xi_1) + L(x,\xi_2).$$

The length of a curve $\gamma$ in the Finsler metric equals

$$S = l_F(\gamma) = \int_\gamma L(x,\dot{x})\,dt > 0.$$

Geometrically we may think of a convex body $K_x$ containing the point $O = \xi$, in the tangent space of vectors $\{\xi\}$ at the point $x \in \mathbb{R}^n$. The function $L(x,\xi)$ equals 1 if $\xi$ lies on the boundary of $K_x$; then homogeneity condition (12.12) defines the Lagrangian uniquely.

Such metrics are induced, e.g., on a surface by its imbedding into a Banach space of any dimension, in the same way as the Euclidean metric generates the Riemannian metric on a surface in $\mathbb{R}^n$.

The homogeneity condition (12.12) is fulfilled also for the Lagrangian

$$L(x,\dot{x}) = \sqrt{g_{ij}(x)\dot{x}^i\dot{x}^j} + \mu A_i(x)\dot{x}^i,$$

which in general does not satisfy the condition of positivity (it always fails for sufficiently large $\mu$) and the triangle inequality.

For this Lagrangian, the naturally parametrized trajectories satisfy the equation

$$\ddot{x}^m + \Gamma^m_{ij}\dot{x}^i\dot{x}^j + g^{km}\mu F_{ik}\dot{x}^i = 0,$$

where

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}.$$

These are equations of the motion of a charged particle with kinetic energy $|\dot{x}|^2/2$ in the magnetic field $F = \sum_{i<k} F_{ik}\,dx^i \wedge dx^k$ (for more details see 12.3.3).

According to Maxwell's equations the magnetic field is specified by a closed form $F$ ($dF = 0$). There are situations when on some topologically nontrivial manifolds (e.g., on $U = \mathbb{R}^3 \setminus \{0\}$) this form is not exact and its vector-potential $A_k\,dx^k$ is defined only locally:

$$F = \sum_{i<j} F_{ij}\,dx^i \wedge dx^j = d(A_k\,dx^k).$$

We will refer to this setting as the *Dirac monopole* and discuss it later on.

## 12.3. Groups of symmetries and conservation laws

**12.3.1. Conservation laws of energy and momentum.** It follows from the Euler–Lagrange equations that if the Lagrangian $L(x, \dot{x})$ does not depend on time $t$ explicitly, then the *energy conservation law* holds: the total derivative of the energy $E$ along an extremal is equal to zero, i.e.,

$$\begin{aligned}\frac{dE}{dt} &= \frac{d}{dt}\Big(\dot{x}^i \frac{\partial L}{\partial \dot{x}^i} - L\Big) = \ddot{x}^i \frac{\partial L}{\partial \dot{x}^i} + \dot{x}^i \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}^i}\Big) - \frac{\partial L}{\partial x^i}\dot{x}^i - \frac{\partial L}{\partial \dot{x}^i}\ddot{x}^i \\ &= \dot{x}^i\Big(\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i}\Big) = 0.\end{aligned}$$

If the Lagrangian $L(x, \dot{x}, t)$ does not depend on the coordinate $x^i$, then the corresponding momentum is conserved (the *momentum conservation law*):

$$\frac{dp_i}{dt} = \frac{\partial L}{\partial x^i} = 0.$$

In this case the coordinate $x^i$ is said to be *cyclic*.

For example, in the case of geodesics, the Lagrangian $L = \frac{1}{2}|\dot{x}|^2$ does not depend on time explicitly, coincides with the energy $E$, and is preserved along the trajectories. Therefore, all the geodesics are naturally parametrized.

The knowledge of conservation laws enables us to simplify the equations of motion and sometimes integrate them completely.

EXAMPLE. GEODESIC FLOW ON A SURFACE OF REVOLUTION. Let a surface in three-dimensional space be specified in cylindrical coordinates $r, \varphi, z$ by the equation

$$r = r(z).$$

Take $z$ and $\varphi$ to be local coordinates on the surface. Since the Euclidean metric in cylindrical coordinates is given by

$$dl^2 = dz^2 + dr^2 + r^2\, d\varphi^2,$$

the first fundamental form of the surface is

$$dl^2 = g_{zz}\, dz^2 + r^2(z)\, d\varphi^2, \quad g_{zz} = 1 + r_z^2.$$

The Lagrangian for the equation of geodesics is equal to

$$L = \frac{1}{2}\,(g_{zz}\dot{z}^2 + r^2(z)\dot{\varphi}^2) = \frac{1}{2}\,(g_{rr}\dot{r}^2 + r^2\dot{\varphi}^2),$$

with energy $E$ equal to $L$ and the momentum corresponding to the cyclic coordinate $\varphi$ equal to

$$p_\varphi = \frac{\partial L}{\partial \dot{\varphi}} = r^2\dot{\varphi}.$$

Both quantities $E$ and $p_\varphi$ are preserved along the trajectories.

Denote by $\alpha$ the angle between the velocity vector $v$ of the geodesic and the tangent vector $e_\varphi$. Then

$$\cos\alpha = \frac{\langle v, e_\varphi\rangle}{\sqrt{\langle v, v\rangle\langle e_\varphi, e_\varphi\rangle}} = \frac{r^2\dot{\varphi}}{\sqrt{2E}r} = \frac{p_\varphi}{\sqrt{2E}r}.$$

This implies that the quantity $r\cos\alpha = \frac{p_\varphi}{\sqrt{2E}}$ (*Clairaut's integral*) is preserved along the trajectories.

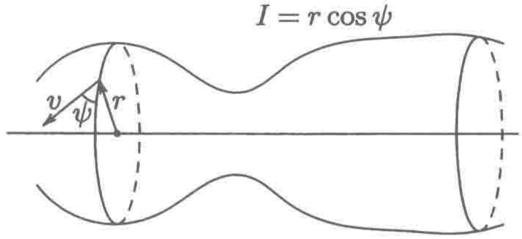


**Figure 12.12.** Clairaut's integral.

**Theorem 12.16.** *The angular component $r\cos\alpha$ of the momentum is preserved along a geodesic on a surface of revolution in $\mathbb{R}^3$.*

Now that we know two independent variables $E$ and $p_\varphi$ that remain unchanged, we can completely integrate the equations of geodesics on a surface of revolution:

$$\frac{dr}{dt} = \sqrt{\frac{2E - p_\varphi^2/r^2}{g_{rr}}}, \quad \frac{d\varphi}{dt} = \frac{p_\varphi}{r^2}$$

(here $g_{rr}$ is locally a function of $r$).

**12.3.2. Fields of symmetries.** Consider the conservation law of momentum from a more general point of view.

We say that a *local one-parameter group* of transformations $S_\tau$, $-\infty < \tau < \infty$, is given in $\mathbb{R}^n$ if for any point of $\mathbb{R}^n$ there is a number $\tau_0 > 0$ and a neighborhood $U$ of this point where the transformation $S_\tau$ is defined and smooth for $|\tau| < \tau_0$:

$$S_\tau : U \to \mathbb{R}^n.$$

Moreover, $S_0 = 1$ is the identity transformation, and the local group properties

$$S_{\tau_1+\tau_2} = S_{\tau_1} \circ S_{\tau_2}, \quad S_{-\tau} = S_\tau^{-1}$$

hold whenever these mappings are defined.

A local one-parameter group is generated by translations along trajectories of a vector field by time $\tau$:

$$X = \frac{d}{d\tau} S_\tau(x)\Big|_{\tau=0}.$$

By the theorem on the existence and uniqueness of solution to an ordinary differential equation, any smooth vector field determines a group of translations along the integral curves of the field $S_\tau$ (see 8.3.1). Moreover, in a small neighborhood of any point $x \in \mathbb{R}^n$ where the vector $X$ is not equal to zero, the vector field can be linearized, namely, there exist special local coordinates $y^1, \dots, y^n$ in which $X = (1, 0, \dots, 0)$, and hence the transformation $S_\tau$ is a translation along the first coordinate:

$$S_\tau(y^1, \dots, y^n) = (y^1 + \tau, y^2, \dots, y^n).$$

A local one-parameter group of transformations $S_\tau$ preserves the Lagrangian $L(x, \xi, t)$ if this Lagrangian is preserved under translations along the trajectories of the group:

$$\frac{d}{d\tau} L(S_\tau(x), S_{\tau *}\xi, t) = 0,$$

where $S_{\tau *}$ is the mapping of the tangent spaces (see 8.2.1). In the special coordinates in which the group is the translation along the first coordinate, this coordinate is cyclic, i.e., the Lagrangian does not depend on it.

This condition is equivalent to the property that the Lie derivative of the function $L(x, \xi, t)$ along the field $X^i$ vanishes:

$$\left.\frac{dL}{d\tau}\right|_{\tau=0} = X^i \frac{\partial L}{\partial x^i} + \frac{\partial X^i}{\partial x^j}\, \xi^j \, \frac{\partial L}{\partial \xi^i} = 0$$

(the field $X$ is a *symmetry field*).

**Theorem 12.17.** *For a one-parameter group of transformations $S_\tau$ preserving the Lagrangian $L$, the momentum component along the field $X = \frac{d}{dt} S_\tau(x)\big|_{\tau=0}$ is conserved:*

$$\frac{d}{dt}\left(X^i \frac{\partial L}{\partial \dot{x}^i}\right) = \frac{d}{dt}\,(X^i p_i) = 0.$$

*In the special coordinates this is the first component of the momentum.*

**Proof.** Suppose the field $X$ is nonzero at the point $x \in \mathbb{R}^n$. Choose special coordinates $(y^1, \dots, y^n)$ near $x$ in which the transformation $S_\tau$ is the translation by $\tau$ along the first coordinate. By the Euler–Lagrange equations,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}^1}\right) = \frac{\partial L(y, \dot{y})}{\partial y^1} \equiv 0,$$

and by the construction of special coordinates, $\frac{\partial L}{\partial \dot{y}^1} = X^i \frac{\partial L}{\partial \dot{x}^i}$. Hence the theorem. $\square$

**12.3.3. Conservation laws in relativistic mechanics.** The motion of a relativistic free particle of nonzero mass $m > 0$ in the Minkowski space with coordinates $(x^0, x^1, x^2, x^3)$, $x^0 = ct$, is determined by one of the two functionals (actions), which are considered only for time-like curves:

$$S_1 = \frac{mc}{2} \int \langle \dot{x}, \dot{x} \rangle \, d\tau, \quad \langle \dot{x}, \dot{x} \rangle = (\dot{x}^0)^2 - \sum_{\alpha=1}^{3} (\dot{x}^\alpha)^2;$$

$$S_2 = -mcl = -mc \int \sqrt{\langle \dot{x}, \dot{x} \rangle} \, d\tau = -mc \int dl.$$

According to the *relativistic principle of least action*, the world lines of free particles with mass are extremals of the functionals $S_1$ or $S_2$.

Similarly to the case of Riemannian metric (see Section 12.2) it is proved that the extremals of these functionals coincide and are time-like curves. The action $S_2$ is better suited for comparison with classical mechanics, which is obtained in the relativistic limit as $v/c \to 0$.

Since the value of the functional $S_2$ does not depend on parametrization of the curve, the parameter may be taken arbitrarily. Usually one takes it to be $\tau = t = x^0/c$. For this parametrization,

$$S_2 = -mcl = -mc^2 \int \sqrt{1 - \left(\frac{v}{c}\right)^2} \, dt,$$

where $v$ is the three-dimensional velocity and $l/c$ is the proper time of the particle. Note that the Lagrangian

$$L = -mc^2 \sqrt{1 - \left(\frac{v}{c}\right)^2}$$

is written in the three-dimensional form. According to the general rules (see 12.2.1), the energy and momentum of such a Lagrangian have the form

$$E = pv - L = \dot{x}^\alpha \frac{\partial L}{\partial \dot{x}^\alpha} - L = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}, \qquad p_\alpha = \frac{mv^\alpha}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

As $v/c \to 0$, we have

$$E = mc^2 \left(1 + \frac{v^2}{2c^2} + \cdots \right), \qquad p_\alpha = mv^\alpha (1 + \cdots),$$

i.e., as a first approximation for the momentum we obtain the classical expression

$$p_\alpha \approx mv^\alpha,$$

and for the energy, the classical expression up to the constant $mc^2$:

$$E \approx mc^2 + \frac{mv^2}{2}.$$

Moreover, the following identity holds:

$$E^2 - c^2p^2 = m^2c^4, \quad E = c\sqrt{p^2 + m^2c^2}.$$

For $E > 0$, the points $(E, cp)$ run over a three-dimensional Lobachevsky space (the mass surface) in the (Minkowski) space of 4-momenta $\mathbb{R}^{1,3}$ with coordinates $(E, cp_1, cp_2, cp_3)$.

For particles of zero mass ($m = 0$), the mass surface degenerates into a cone, with positive part of the form $E = c|p|$. This relation determines the law of motion of photons by means of the Hamilton equations.

In terms of this formalism the time is a special coordinate. So far, special relativity looks simply like usual mechanics with somewhat different kinetic energy.

These conceptions are appropriate if we consider the motion of a particle in an exterior field provided the particle does not affect the field (the influence is negligible).

Given an electromagnetic field $F_{\mu\nu}$, $\nu, \mu = 0, 1, 2, 3$, the Lagrangian for a particle with mass $m > 0$ and charge $e$ is

$$L_2\, dt = -mc^2\sqrt{1 - \left(\frac{v}{c}\right)^2}\, dt + \frac{e}{c}\, A_\alpha(x(t))\, dx^\alpha - eA_0\, dt.$$

Here $dx^\alpha/dt = v^\alpha$, $\alpha = 1, 2, 3$, $c$ is the velocity of light in vacuum, and $A_\alpha$ is the vector-potential of the electromagnetic field,

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}, \quad A_0 = U.$$

Recall that $F_{0\alpha} = E_\alpha$ is the electric field, and $F_{\alpha\beta}$ is the magnetic field, $\alpha, \beta = 1, 2, 3$ (see 9.1.3).

Now we state the fundamental *principle of inclusion of the electromagnetic field.*

Both in classical mechanics and special or general relativity, inclusion of the electromagnetic field is done as follows.

Suppose we know the Lagrangian $L(x, \dot{x})$ of a particle with charge $e$ when there is no field. Then the action of the particle in the presence of the field is

$$S = \int L(x, \dot{x})\, dt + \frac{e}{c} \int A_\alpha\, dx^\alpha,$$

where the integral is taken along the trajectory. Here $A_\alpha\,dx^\alpha$ is the 1-form (4-covector) called the vector-potential of the electromagnetic field, and so

$$d(A_\alpha\,dx^\alpha) = \sum_{\mu<\nu} F_{\mu\nu}\,dx^\mu \wedge dx^\nu,$$
$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}, \quad \mu,\nu = 0,1,2,3,$$
$$x^0 = ct.$$

The vector-potential is defined up to addition of the gradient,

$$A_\alpha\,dx^\alpha \to \left(A_\alpha + \frac{\partial\varphi}{\partial x^\alpha}\right)dx^\alpha,$$

which does not affect the equation of extremals.

The vector-potential is often "calibrated" (fixed) by the requirement

$$A_0 = 0.$$

Then

$$F_{\alpha\beta} = \frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta} \quad \text{(magnetic field)},$$
$$F_{0\alpha} = E_\alpha = \frac{\partial A_\alpha}{\partial x^0} \quad \text{(electric field)}.$$

Now we return to the functional

$$S_1 = \frac{mc}{2}\int \left\langle \frac{dx}{d\tau}, \frac{dx}{d\tau}\right\rangle d\tau$$

in order to clarify the 4-dimensional meaning of relativistic mechanics. For a free particle with mass ($m > 0$) the extremal is naturally parametrized (by the conservation law of the formal "energy"). Define the 4-dimensional momentum vector $\tilde{p}_i = (\tilde{p}_0, \tilde{p}_\alpha)$, $\alpha = 1,2,3$, as

$$\tilde{p}_0 = \frac{\partial L}{\partial x'^i}, \quad i = 0,1,2,3, \quad \text{where } x'^i = \frac{dx^i}{d\tau}.$$

We obtain

$$\tilde{p}_0 = \frac{\partial L}{\partial x'^0} = mcx'^0, \quad \tilde{p}_\alpha = \frac{\partial L}{\partial x'^\alpha} = -mcx'^\alpha, \quad \alpha = 1,2,3.$$

Upon raising the index in the Minkowski metric we obtain the vector

$$\tilde{p}^i = mcx'^i, \quad i = 0,1,2,3.$$

Since the parameter on the extremals of the functional $S_1$ is natural, we have

$$dl = \sqrt{1 - \frac{v^2}{c^2}}\, dt, \quad x'^i = \frac{dx^i}{dl},$$
$$\tilde{p}^0 = mcx'^0 = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} = E,$$
$$\tilde{p}^\alpha = mcx'^\alpha = mc\frac{dx^\alpha}{dt}\frac{1}{\sqrt{1 - \frac{w^2}{c^2}}} = \frac{mcv^\alpha}{\sqrt{1 - \frac{v^2}{c^2}}} = cp^\alpha.$$

Therefore, the 4-*vector of energy-momentum* (or the *momentum* 4-*vector*) $\tilde{p}^i$ is related to the three-dimensional momentum and energy by the equalities

$$\tilde{p}^0 = E, \quad \tilde{p}^\alpha = cp^\alpha.$$

We see that the energy-momentum vector $(E, cp)$ transforms as a 4-vector under Lorentz transformations, and the 4-dimensional momentum vectors of particles with mass lie on the mass surface endowed with Lobachevsky geometry:

$$E^2 - c^2p^2 = (\tilde{p}^0)^2 - \sum_{\alpha=1}^{3} (\tilde{p}^\alpha)^2 = m^2c^4.$$

For example, under transition to a coordinate system moving uniformly with velocity $v$ along the $x^1$-axis, we have

$$(E, cp) \to (E', cp'),$$

where

$$E' = \frac{E - pv}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad p_1' = \frac{p_1 - \frac{Ev}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad p_2' = p_2, \quad p_3' = p_3.$$

**12.3.4. Conservation laws in classical mechanics.** Consider a system of $n$ classical particles in $\mathbb{R}^3$ with pairwise interaction. It is described by a Lagrangian in $\mathbb{R}^{3n}$ of the form

$$L = \sum_{i=1}^{n} \frac{m_i \dot{x}_i^2}{2} - U(x_1, \dots, x_n), \quad x_i = (x_i^1, x_i^2, x_i^3),$$

where the $x_i$ are the coordinates of the particles, the $m_i$ are their masses, and $U = \frac{1}{2}\sum_{i \neq j} V(x_i, x_j)$ is the interaction potential.

We will require that the system be translation invariant, i.e., that the Lagrangian remain unchanged under translations by vectors $\xi \in \mathbb{R}^3$:

$$x_i \to x_i + \xi, \quad \xi = (\xi^1, \xi^2, \xi^3).$$

This holds, e.g., if the potential of pairwise interaction depends only on the difference of the arguments:

$$V(x_i, x_j) = V(x_i - x_j).$$

**Theorem 12.18.** *For a translation invariant system of particles with pairwise interaction, the total momentum is conserved:*

$$P = \sum_{i=1}^{n} m_i \dot{x}_i, \quad \frac{dP}{dt} = 0.$$

**Proof.** There are three groups of symmetries acting in $\mathbb{R}^{3n}$:

$$S_\tau^\alpha: \quad x_i^\alpha \to x_i^\alpha + \tau, \quad x_i^\beta \to x_i^\beta \text{ for } \beta \neq \alpha,$$

where $\alpha, \beta = 1, 2, 3$. By Theorem 12.17, corresponding to these groups are three quantities $P_1, P_2, P_3$, the components of the total momentum, which are conserved. Hence the theorem. □

EXAMPLE. SYSTEM OF TWO PARTICLES. Let $n = 2$ and let

$$L = m_1|\dot{x}_1|^2 + m_2|\dot{x}_2|^2 - 2V(x_1 - x_2).$$

We pass to a uniformly moving coordinate system in which $P = 0$. Then

$$m_1\dot{x}_1 + m_2\dot{x}_2 = 0.$$

We place the origin at the center of mass. Then

$$m_1 x_1 + m_2 x_2 = 0, \quad V(x_1 - x_2) = V\Big(x_1 + \frac{m_1}{m_2} x_1\Big).$$

Set $m^* = \frac{m_1 m_2}{m_1 + m_2}$ and $U(x_1) = V\big((1 + \frac{m_1}{m_2})x_1\big)$. Then the Newton equations

$$m_1 \ddot{x}_1^\alpha = -\frac{\partial V(x_1 - x_2)}{\partial x_1^\alpha}, \quad \alpha = 1, 2, 3,$$

reduce to the equations

$$m^* \ddot{x}_1^\alpha = -\frac{\partial U(x_1)}{\partial x_1^\alpha}.$$

Hence we obtain the following theorem.

**Theorem 12.19.** *The problem concerning the motion of two particles with translation-invariant interaction potential in the coordinate system with origin at the mass center is equivalent to the problem of the motion of a single particle with reduced mass $m^* = \frac{m_1 m_2}{m_1 + m_2}$ in the field with potential $U(x_1) = V(x_1 - x_2)$, where $m_1 x_1 + m_2 x_2 = 0$.*

Now we consider systems that are invariant relative to the group of rotations SO(3).

If the Lagrangian of a particle in $\mathbb{R}^3$ is invariant under all rotations, it is said to be *spherically symmetric*. To the one-parameter subgroups of rotations about the $x$-, $y$-, and $z$-axes correspond their generators $L_x, L_y, L_z$:

$$L_x = (0, -z, y), \quad L_y = (z, 0, -x), \quad L_z = (-y, x, 0)$$

(see 8.3.3). They give rise to the conservation laws for spherically symmetric Lagrangians:

$$M_x = L_x^\alpha p_\alpha, \quad M_y = L_y^\alpha p_\alpha, \quad M_z = L_z^\alpha p_\alpha,$$
$$\frac{d}{dt} M_x = \frac{d}{dt} M_y = \frac{d}{dt} M_z = 0,$$

which can be rewritten as

$$M_x = yp_z - zp_y, \quad M_y = zp_x - xp_z, \quad M_z = xp_y - yp_x. \tag{12.13}$$

Thus we have proved the following theorem.

**Theorem 12.20.** *The motion of a particle in $\mathbb{R}^3$ with a spherically symmetric Lagrangian conserves the vector*

$$M = (M_x, M_y, M_z) = [x, p],$$

*where $x = (x, y, z)$ and $p$ is the momentum of the particle.*

This vector $M = [x, p]$ is called the *angular momentum*.

In $\mathbb{R}^n$ the angular momentum has the form

$$M = (M_{ij}) = x^i p_i - x^j p_j = \vec{x} \wedge \vec{p}$$

(exterior product); hence it is in fact a skew-symmetric 2-tensor rather than a vector.

EXAMPLE. COMPLETE SOLUTION OF THE TWO-PARTICLE PROBLEM. Let

$$L = m_1 |\dot{x}_1|^2 + m_2 |\dot{x}_2|^2 - U(|x_1 - x_2|).$$

Since the potential depends on the distance between the particles, $r = |x_1 - x_2|$, this system is invariant relative to the entire group of motions of $\mathbb{R}^3$. Using Theorem 12.19 (invariance relative to translations), we pass to the problem for a single particle of mass $m$ in the field $U(r)$. For such a particle, the conservation laws of the angular momentum $M = [x, p]$ and the energy

$$E = pv - L = \frac{mv^2}{2} + U(r), \quad v = \dot{x},$$

hold, since the Lagrangian does not depend on time.

**Lemma 12.17.** *The particle moves in the plane spanned by the vectors $x$ and $p$.*

**Proof.** Since the angular momentum $M$ is conserved and $\dot{x} = p/m$, the vector $[x, \dot{x}] = M/m$ has constant direction, which is orthogonal to the plane $(x, p)$. Hence the lemma. $\square$

We pass to cylindrical coordinates $(z, r, \varphi)$ with $z$-axis in the direction of the vector $M$. We obtain

$$L = \frac{m\dot{x}^2}{2} - U(r) = m\Big(\frac{\dot{r}^2 + r^2\dot{\varphi}^2}{2}\Big) - U(r),$$

$$p_\varphi = M = mr^2\dot{\varphi} = \frac{\partial L}{\partial \dot{\varphi}} = \text{const}, \quad \dot{\varphi} = \frac{M}{mr^2},$$

$$E = \frac{m}{2}\Big(\dot{r}^2 + \frac{M^2}{m^2r^2}\Big) + U(r) = \frac{m\dot{r}^2}{2} + U_{\text{eff}}(r),$$

where

$$U_{\text{eff}}(r) = U(r) + \frac{M^2}{2mr^2}.$$

The problem on the motion of the particle reduces to the one-dimensional problem (in $r$) with potential $U_{\text{eff}}(r)$, which is solved explicitly by the formulas

$$\dot{r}^2 = \frac{2}{m}(E - U_{\text{eff}}),$$

$$t - t_0 = \int \frac{dr}{\sqrt{\frac{2}{m}(E - U_{\text{eff}})}},$$

$$\varphi - \varphi_0 = \int \frac{M\,dt}{mr^2}.$$

Excluding $t$, one can obtain the equation of the orbit $\varphi = \varphi(r)$ or $r = r(\varphi)$.

In the two important well-known cases, $U = \alpha/r$ (Newton's potential of the gravitation field) and $U = \alpha r^2$, there are entire domains in the $(x, \dot{x})$-space filled out by closed orbits:

$$\text{the domain } \ E < 0 \ \text{ for } \ U = \frac{\alpha}{r} \ \text{ and } \ \alpha < 0,$$

$$\text{the domain } \ E \geq 0 \ \text{ for } \ U = \alpha r^2 \ \text{ and } \ \alpha > 0.$$

The domain $E < 0$ in the case $U = \alpha/r$ is the domain of Kepler's ellipses. For the orbit to be closed, the following nontrivial equality is required:

$$r(\varphi + 2\pi n) \equiv r(\varphi), \tag{12.14}$$

where $n$ is an integer.

For analytic spherically symmetric potentials $U(r) \neq \alpha/r$, $\alpha r^2$, equality (12.14) cannot be fulfilled in an entire domain of the phase space, and in general the closed orbits fill out only a set of measure zero.

Therefore, Newton's potential $U(r) = \frac{\alpha}{r}$ has a deep latent symmetry, which disappears under any small perturbation. This is a fact of fundamental significance.

**12.3.5. Systems of relativistic particles and scattering.** Let $L(x, \dot{x})$, where $x = (x^0, x^1, x^2, x^3)$, be a Lagrangian in the Minkowski space $\mathbb{R}^{1,3}$ invariant relative to the group $\mathrm{O}(1,3)$. Linear vector fields that determine one-parameter subgroups have the form

$$X^i(x) = x_k A^{ki} = g_{kl} x^l A^{ki},$$

where $g_{kl}$ is the Minkowski metric and $A^{ki}$ is an arbitrary skew-symmetric matrix (see 8.3.3). To any such vector field $X$ corresponds the conservation law of the $X$-component of the 4-momentum,

$$p_i X^i = p_i x_k A^{ki} = \text{const},$$

where $p_i = \frac{\partial L}{\partial \dot{x}^i}$. Because of skew-symmetry of the matrix $A^{ki}$ we have

$$p_i x_k A^{ki} = \frac{1}{2}(p_i x_k - p_k x_i) A^{ki} = \frac{1}{2} M_{ik} A^{ki} = \text{const}.$$

Since the matrix $A^{ki}$ may be chosen arbitrarily, all the components of the tensor

$$M_{ik} = x_i p_k - x_k p_i,$$

which is called the *momentum 4-tensor*, are conserved.

For the corresponding tensor $M^{ik}$ with superscripts, the spacial components $M^{\alpha\beta}$, $\alpha, \beta = 1, 2, 3$, are

$$M^{\alpha\beta} = (x^\alpha p^\beta - x^\beta p^\alpha),$$

and hence they are equal to the components of the three-dimensional momentum vector

$$M = [x, p],$$
$$M^{23} = M_x, \quad M^{31} = M_y, \quad M^{12} = M_z.$$

The components $M^{01}$, $M^{02}$, $M^{03}$ form a three-dimensional vector

$$ctp - \frac{E}{c} x = (M^{01}, M^{02}, M^{03}). \tag{12.15}$$

Consider now a system of $n$ relativistic particles $x_1, \ldots, x_n$ with Lagrangian $L(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)$, which is invariant relative to the Poincaré group, i.e., the group of motions of the Minkowski space $\mathbb{R}^{1,3}$. We will require

the motion to act on all the particles $x_i$. We have the following conservation laws:

$$\left(\sum_{i=1}^{n} \frac{E_i}{c}, \sum_{i=1}^{n} p_i\right) = \text{const} \quad \text{(conservation of the total momentum 4-vector)},$$

$$\sum_{i=1}^{n} M_i^{kl} = \text{const} \quad \text{(conservation of the total momentum tensor)},$$

where $M_i^{kl}$ is the momentum tensor of the $i$th particle. Now formula (12.15) implies that

$$\frac{\sum E_i x_i}{\sum E_i} = t\,\frac{\sum c^2 p_i}{\sum E_i} + \text{const}\,.$$

Therefore, the point

$$x = \frac{\sum E_i x_i}{\sum E_i}, \tag{12.16}$$

which is the relativistic analog of the mass center, moves with constant velocity

$$v = \frac{\sum c^2 p_i}{\sum E_i}.$$

In the limit, as the velocities of particles are small relative to $c$, we have $E_i \approx m_i c^2$ and formula (12.16) turns into the classical formula for the mass center

$$x = \frac{\sum m_i x_i}{\sum m_i}.$$

Note that the relativistic mass center is not invariant with respect to the choice of coordinate system.

In the physics of real elementary particles and high energies, the classical treatment without recourse to quantum theory is possible only when the particles are already far away from each other. We assume that as $t \to -\infty$, there were $m$ incoming particles, and as $t \to +\infty$, there are $n$ outgoing product-particles. Nevertheless, the total conservation laws corresponding to the Poincaré group are considered unquestionable whatever the interaction in the interim, provided all the incoming and outgoing particles as $t \to \pm\infty$ are taken into account.

For example, we point out that it is impossible for a process to have two particles of mass $m_i > 0$ as $t \to -\infty$ to become a single particle of mass $m = 0$ as $t \to +\infty$. We leave it as a problem for the reader to prove that this would contradict the conservation laws stated above.

Consider an elastic process with two massive particles, i.e., a process where the masses remain unchanged and no new particles appear. The

physical assumptions about the process are stated relative to the mass center system: the total 4-momentum is equal to zero. Therefore, for the 3-momenta we have $p_{\pm}^{(1)} = -p_{\pm}^{(2)}$ as $t \to \pm\infty$, as well as $|p_{\pm}^{(i)}| = |p_{\pm}^{(i)}|$, with all of them lying in the same plane. Consequently, the vector $p_{+}^{(i)}$ is obtained from $p_{-}^{(i)}$ by a planar rotation through an angle $\varphi$. Experimental measurements are performed in another frame where one of the particles (heavy) stays at rest (the 4-momentum is $P = (Mc^2, 0, 0, 0)$), while the light particle hits it with 4-momentum $p = (\epsilon, p_1, p_2, p_3)$, $\epsilon^2 - c^2 \sum p_i^2 = m^2c^4$. Applying the Lorentz transformation first from the experimental frame as $t \to -\infty$ and then conversely as $t \to +\infty$, when the process is completed, one can find the result of elastic scattering as a function of the angle $\varphi$. We also leave it as a problem.

## 12.4. Hamilton's variational principle

**12.4.1. Hamilton's theorem.** A Lagrangian $L(x, \dot{x})$ is said to be *nondegenerate* if

$$\det\left(\frac{\partial^2 L}{\partial \dot{x}^\alpha \partial \dot{x}^\beta}\right) \neq 0.$$

By the inverse function theorem, for a nondegenerate Lagrangian the mapping

$$\dot{x} \to p = \frac{\partial L}{\partial \dot{x}}$$

of tangent vectors to covectors at a given point $x$ is locally invertible. A Lagrangian is said to be *strongly nondegenerate* if this mapping is invertible globally, i.e., there exist smooth functions $\dot{x}^i = v^i(x, p)$ for all $x, \dot{x}$.

For example, the Lagrangian $L = g_{ij}\dot{x}^i\dot{x}^j$, where $g_{ij}$ is a Riemannian metric, is strongly nondegenerate, and the velocities are related to the momenta by linear transformations

$$p_i = g_{ij}x^j, \quad \dot{x}^i = g^{ij}p_j.$$

Assume that the Lagrangian is strongly nondegenerate. Let us introduce the basic definitions.

The energy $E = \dot{x}^a L_{\dot{x}^a} - L$ expressed as a function of $x$ and $p = L_{\dot{x}}$ is called the *Hamilton function* or *Hamiltonian.*

The transformation

$$L(x, v) \to H(x, p) = v^i \frac{\partial L}{\partial v^i} - L$$

is called the *Legendre transformation.*

The space $\mathbb{R}^n$ of all possible states $x \in \mathbb{R}^n$ of a system is called the *configuration space*, and the space $\mathbb{R}^{2n}$ with coordinates $(x, p)$ is called the *phase space.*

The variational problem was initially posed for trajectories in the tangent bundle $TM^n = T\mathbb{R}^n$ to the configuration space. Under the Legendre transformation, the Euler–Lagrange equations go into the Hamilton equations.

**Theorem 12.21** (Hamilton). *Let $L(x, \dot{x})$ be a nondegenerate Lagrangian, and let $H(x,p)$ be the corresponding Hamiltonian, where $p = \frac{\partial L(x,\dot{x})}{\partial \dot{x}}$ and $\dot{x} = v(x,p)$ is the local inversion of the mapping $\dot{x} \to p$. The Euler–Lagrange equations*

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = 0$$

*are locally equivalent to the* Hamilton equations *in which $x$ and $p$ are treated as independent variables:*

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x} \tag{12.17}$$

*in any domain of the phase space $(x,p)$ where the Legendre transformation is one-to-one.*

**Proof.** Let $v = \dot{x}$ and

$$H(x,p) = pv - L(x, v(x,p)).$$

Since $p = \frac{\partial L}{\partial v}$, the Euler–Lagrange equations imply the equalities

$$\frac{\partial H}{\partial p} = \frac{\partial}{\partial p}(pv - L) = v + p\frac{\partial v}{\partial p} - \frac{\partial L}{\partial v}\frac{\partial v}{\partial p} = v,$$

$$-\frac{\partial H}{\partial x} = -\frac{\partial}{\partial x}(pv - L) = -p\frac{\partial v}{\partial x} + \frac{\partial L}{\partial x} + \frac{\partial L}{\partial v}\frac{\partial v}{\partial x} = \dot{p},$$

which constitute the Hamilton equations (12.17). Hence the theorem. □

Rewrite the action $S = \int L\,dt$ in terms of the phase space:

$$S = \int L\,dt = \int (p\dot{x} - H(x,p))\,dt$$

and consider it for all curves in the phase space without assuming that $\dot{x} = v(x,p)$. This action is related to the variational problem for the curves $(x(t), p(t))$ in $\mathbb{R}^{2n}$; the following fundamental theorem (from the viewpoint of modern symplectic geometry) holds for this problem.

**Theorem 12.22** (Hamilton's variational principle). *The equations of extremals for the functional*

$$S = \int (p\dot{x} - H(x,p))\,dt \tag{12.18}$$

*in the $2n$-dimensional phase space with coordinates $(x,p)$ coincide with the Hamilton equations and do not have any new solutions. This functional has no nontrivial minima.*

**Proof.** Let $(z^1, \dots, z^{2n})$ be the coordinates $(x^1, \dots, x^n, p_1, \dots, p_n)$ in $\mathbb{R}^{2n}$ and let $\tilde{L}(z, \dot{z}) = p\dot{x} - H(x,p)$. The Euler–Lagrange equations for this Lagrangian have the form

$$\dot{p}_i = \frac{d}{dt}\Big(\frac{\partial \tilde{L}}{\partial \dot{x}^i}\Big) = \frac{\partial \tilde{L}}{\partial x^i} = -\frac{\partial H}{\partial x},$$
$$0 = \frac{d}{dt}\Big(\frac{\partial \tilde{L}}{\partial \dot{p}_i}\Big) = \frac{\partial \tilde{L}}{\partial p_i} = \dot{x}^i - \frac{\partial H}{\partial p_i}.$$

Hence the theorem. □

**12.4.2. Lagrangians and time-dependent changes of coordinates.** Consider Lagrangians $L(x, \dot{x}, t)$ of the most general form, which explicitly depend on time.

Under a change of coordinates $x = x(x', t)$ that explicitly depends on time, the action $S = \int L\, dt$ remains unchanged. Hence the Lagrangian $L$ may change only by a total derivative:

$$L(x, \dot{x}, t) \to L(x(x', t), \dot{x}, t) + \frac{df(x,t)}{dt}.$$

If the change of coordinates does not explicitly depend on time, $x = x(x')$, then

$$\dot{x}^i = \frac{\partial x^i}{\partial x^{i'}}\dot{x}^{i'}, \quad \tilde{L}(x', \dot{x}', t) = L\Big(x(x'), \frac{\partial x}{\partial x'}\,\dot{x}', t\Big),$$
$$p_i' = \frac{\partial \tilde{L}}{\partial \dot{x}'^i} = \frac{\partial L}{\partial \dot{x}^i}\,\frac{\partial x^i}{\partial x^{i'}} = p_i\,\frac{\partial x^i}{\partial x'^i},$$

i.e., the velocity is a vector, the momentum is a covector, and the Lagrangian is a scalar. In this case, the energy remains unchanged,

$$E = p_i\dot{x}^i - L = p_i'\dot{x}'^i - \tilde{L} = E'.$$

Now we proceed to coordinate changes depending on time:

$$x = x(x', t), \quad t = t'.$$

Having fixed a time instant $t_0$, we assume that the change is "instantaneous". This means that at time $t = t_0$ we have $x' = x$, i.e., at $t = t_0$, the coordinate systems $x$ and $x'$ coincide.

There is a general formula

$$\dot{x}^i = \frac{\partial x^i}{\partial x^{i'}}\,\dot{x}^{i'} + \frac{\partial x^i}{\partial t} = \frac{\partial x^i}{\partial x^{i'}}\,\dot{x}^{i'} + a^i(x', t),$$

which in the case of instantaneous change at $t = t_0$ becomes

$$\dot{x}^i = \dot{x}'^i + a^i(x', t).$$

Now we pass to a moving coordinate system $x'$:

$$L \to L' = L, \quad L(x, v, t) = L(x', v' + a, t);$$
$$v \to v' = v - a(x', t);$$
$$p \to p' = p, \quad \text{since } \frac{\partial L}{\partial v} = \frac{\partial L'}{\partial v'};$$
$$E \to E' = p'v' - L' = p_i(v^i - a^i) - L = E - p_i a^i(x', t).$$

We see that for $t = t_0$ the momentum does not change and the Hamiltonians undergo translation:

$$H(x, p) \to H(x', p') - p'_i a^i(x', t) = H'(x', p', t), \quad a = \left.\frac{\partial x}{\partial t}\right|_{t=t_0}.$$

Since $x' = x$ and $p' = p$ for $t = t_0$, we obtain

$$\dot{x}' = \frac{\partial H'}{\partial p'} = \frac{\partial H}{\partial p} - a = \dot{x} - a, \quad \dot{p}' = -\frac{\partial H'}{\partial x'} = -\frac{\partial H}{\partial x} = \dot{p}.$$

Thus we have proved the following theorem.

**Theorem 12.23.** *Under the transition to a moving coordinate system $x'$ that coincides with $x = x(x', t)$ at the initial time $t = t_0$, the coordinate and momentum remain unchanged (for $t = t_0$), and the Hamiltonian (energy) changes by $-(p, a)$, where*

$$a = \left.\frac{\partial x}{\partial t}\right|_{t=t_0}, \quad H \to H' = H - (p, a), \quad a = a(x', t).$$

Now we will give examples of application of this theorem.

EXAMPLES. 1. TRANSLATIONAL MOVEMENT OF THE COORDINATE SYSTEM. In this case $a = a(t)$ does not depend on the space point $x'$. Let

$$L = \frac{m\dot{x}^2}{2} - U(x).$$

Then

$$p = mv, \quad p' = p = mv = m(v' + a) = mv' + ma,$$
$$H' = H - (p', a).$$

The Newton equations become

$$m\dot{v} = f = m\dot{v}' + m\dot{a}, \quad m\dot{v}' = f - m\dot{a} = f'.$$

The force $f$ gets the additional inertial term $-m\dot{a}$.

2. ROTATION OF THE COORDINATE SYSTEM IN $\mathbb{R}^3$. Let $a = [\Omega, x']$, where $\Omega$ is the angular velocity (a constant vector). We have

$$H' = H - (p', a) = H - (p', [\Omega, x']),$$
$$p' = p, \quad x' = x \quad \text{for } t = t_0.$$

For the Lagrangian $L = \frac{m\dot{x}^2}{2} - U(x)$ at $t = t_0$ we obtain

$$p = p' = m\dot{x} = m(\dot{x}' + [\Omega, x']) = m\dot{x}' + ma,$$

and the Hamilton equations for $H'$ imply that

$$\dot{p}' = m\ddot{x}' + m[\Omega, \dot{x}']^{\cdot} = -\frac{\partial H}{\partial x} + [(m\dot{x}' + ma), \Omega]$$

(here we take into account that $\dot{\Omega} = 0$). Using the equality $a = [\Omega, x']$, we rewrite the last equation as

$$m\ddot{x}' = f + 2m[\dot{x}', \Omega] + m[[\Omega, x'], \Omega],$$

where $f = -\frac{\partial H}{\partial x}$ (the "old" force in the coordinate system $x$). If $|x'|$ is small or simply bounded, then the last term is of order $|\Omega|^2$ as $\Omega \to 0$. The force $2m[\dot{x}', \Omega]$ is called the *Coriolis force.* Finally we obtain

$$m\ddot{x}' = f + 2m[\dot{x}', \Omega] + O(\Omega^2).$$

3. Again let $L = \frac{m\dot{x}^2}{2} - U(x)$. Then the transition to a moving coordinate system has the form

$$H \to H - p_i a^i, \quad p_i = mv^i = p'_i = m(v'^i + a^i),$$

where

$$H = \frac{p^2}{2m} + U(x),$$

$$H' = \frac{p'^2}{2m} + U(x) - pa = \frac{(p' - ma)^2}{2m} + U(x) - \frac{ma^2}{2}.$$

Therefore, in the moving coordinate system the relationship between the momentum and velocity changes:

$$mv' = p' - ma,$$

and the potential gets the additional term $-\frac{ma^2}{2}$.

4. Inclusion of electromagnetic field. Let $L(x, \dot{x})$ be a Lagrangian. Define a new Lagrangian by the formula

$$\tilde{L} = L + \frac{e}{c} A_i \dot{x}^i,$$

where $A_i$ is the vector-potential of the electromagnetic field and $e$ is the charge. Then the action becomes

$$\tilde{S} = \int \Big( L\,dt + \frac{e}{c}\, A_i\, dx^i \Big).$$

The operation of adding the term $\frac{e}{c} A_i \dot{x}^i$ to the Lagrangian is called "inclusion of the field", as was already pointed out in 12.3.3.

If $H(x, p) = pv - L$ and $\tilde{H}(x, \tilde{p}) = \tilde{p}v - \tilde{L}$, where $\tilde{p} = \frac{\partial \tilde{L}}{\partial v}$, then

$$\tilde{p}_i = p_i + \frac{e}{c} A_i, \quad \tilde{H}(x, \tilde{p}) = H\Big(x, \tilde{p} - \frac{e}{c} A\Big).$$

Thus inclusion of the field is equivalent to a translated momentum in the Hamiltonian and is similar in that to the transition to a moving coordinate system.

It is worthwhile to point out that in the framework of Hamiltonian formalism this operation is applicable also to particles of zero mass (although no charged particles of this kind are known):

$$H(x,p) = c\left|p - \frac{e}{c}\,A(x)\right|.$$

**12.4.3. Variational principles of Fermat type.** Hamilton's variational principle (Theorem 12.22) allows us to obtain the following result.

**Theorem 12.24.** *Let $H(x,p)$ be a Hamiltonian independent of time. Fix an energy level $H = E = \mathrm{const}$ and define the truncated action*

$$S_0 = \int p\dot{x}\,dt = \int p\,dx \tag{12.19}$$

*for curves $(x(t), p(t))$ lying on this level. Then the trajectories of the Hamiltonian system with energy $E$ are precisely the extremals of the functional $S_0$ on the space of all curves $(x(t), p(t))$ with given energy $H(x(t), p(t)) \equiv E$.*

**Proof.** By Theorem 12.22, the initial functional $S = \int(p\,dx - H\,dt)$ attains an extremum on the surface $H(x,p) = E$ on all initial extremals (solutions to the Hamiltonian system). Now we consider the extremal problem on the narrower class of curves such that $H(x,p) = E$. Since

$$S = \int(p\,dx - E\,dt) = \int(p\,dx - d(Et)),$$

the Lagrangians $p\dot{x} = \tilde{L}(z,\dot{z})$ and $\tilde{L}(z,\dot{z}) - \frac{d}{dt}(Et)$ in coordinates $z^1, \dots, z^{2n-1}$ on the surface $H(x,p) = E$ (of dimension $2n-1$) are equivalent. Therefore, all the extremals of the initial functional in the $(x,p)$-space are the extremals of the new (truncated) action $S_0 = \int p\,dx$ on the surface $H = E$. Hence the theorem. □

Making use of the truncated action we can prove the following theorem.

**Theorem 12.25** (Maupertuis–Fermat–Jacobi principle). *If $(x(t), p(t))$ is a solution to a Hamiltonian system with Hamiltonian $H = p^2/2m + U(x) = E$ with fixed energy $E$, then the curve $x(t)$ is a geodesic of the new metric*

$$g_{ij} = 2m(E - U(x))\delta_{ij},$$

*but the parameter $t$ on this curve is not natural relative to this metric.*

**Proof.** Let

$$L = \frac{m\dot{x}^2}{2} - U(x), \quad H(x,p) = \frac{p^2}{2m} + U(x).$$

The following equation holds along the extremals:

$$\dot{x} = \frac{\partial H}{\partial p} = \frac{p}{m}.$$

We introduce an even smaller class of curves on the surface $H = E$ by requiring the relation $\dot{x} = p/m$ to hold on them, and restrict the truncated action $S_0 = \int p\,dx$ to this class. Then the equality $H = E$ implies that

$$|p| = \sqrt{2m(E - U(x))},$$

and since $\dot{x} = p/m$, we have

$$(p, \dot{x}) = p_\alpha \dot{x}^\alpha = |p||\dot{x}|.$$

Hence

$$\begin{aligned} S_0 &= \int p\,dx = \int (p, \dot{x})\,dt = \int |p||\dot{x}|\,dt = \int |p|\,|dx| \\ &= \int \sqrt{2m(E - U(x))}\,|dx|. \end{aligned}$$

This completes the proof. □

Note that this theorem continues to hold in a more general case where

$$L = \frac{1}{2}\, g_{ij}\dot{x}^i\dot{x}^j - U(x).$$

In this case the extremals specify geodesics of the metric $2m(E - U(x))g_{ij}$ (the proof is similar).

Consider now the natural Hamiltonian with included electromagnetic field

$$H = \frac{\left(p - \frac{e}{c}\, A(x)\right)^2}{2m} + U(x),$$

where $A(x) = (A_1, A_2, A_3)$ is the vector-potential of the magnetic field

$$B_{\alpha\beta} = \left(\frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta}\right), \quad \alpha, \beta = 1, 2, 3.$$

Repeating word for word the above argument, we arrive at the following extension of the preceding theorem: The trajectories of the motion of a particle in the magnetic field $B_{\alpha\beta}$ on the energy level $E$ specify the extremals of the functional

$$\tilde{S}_E = \int \sqrt{2m(E - U(x))}\sqrt{\sum_{\alpha=1}^{3}(dx^\alpha)^2} + \frac{c}{e}\int A_\alpha(x)\,dx^\alpha.$$

As usual, the integral is taken here along the trajectory.

It is worth pointing out that it is only the magnetic field tensor $B_{\alpha\beta}$ that has a direct physical meaning. The vector-potential $A_\alpha\,dx^\alpha$ is in general defined only locally by the condition $d^{-1}B = A$, and still not uniquely. We will discuss the invariant meaning of such variational principles later on.

Consider now the Hamiltonian

$$H = c(x) \cdot |p|$$

describing the trajectories of light in an isotropic medium with light variable velocity $c(x)$. Restrict the truncated action $S_0 = \int p\,dx$ to the set of curves such that $H(x,p) = E$ and

$$\dot{x} = \frac{\partial H}{\partial p} = c(x)\frac{p}{|p|}$$

(obviously, $|\dot{x}| = c(x)$). Since $|p| = H/c(x) = E/c(x)$ and $\langle p, dx\rangle = |p| \cdot |dx|$, we have

$$\int \langle p, dx\rangle = \int |p| \cdot |dx| = \int \frac{E}{c(x)}\,|dx| = E\int \frac{|dx|}{c(x)} = E\int \sqrt{\frac{dx^2}{c^2(x)}}.$$

The integral $\int_\gamma \frac{|dx|}{c(x)}$ is, obviously, equal to the time of motion of light along the path $\gamma$. Thus we have proved the following theorem.

**Theorem 12.26** (Fermat's principle). *Light moves along the curve for which the time of motion has extremum among all smooth curves joining given points. For an isotropic medium these curves are geodesics of the metric*

$$g_{ij} = \frac{1}{c^2(x)}\delta_{ij}.$$

## Exercises to Chapter 12

1. Prove that for the mappings $f\colon \mathrm{SO}(n) \to S^{n-1}$ and $g\colon \mathrm{U}(n) \to S^{2n-1}$ which assign to a matrix its first column, all points are regular. Find the inverse images of points under these mappings.

2. Prove that the mappings $f\colon S^n \to S^n$ and $g\colon S^n \to S^n$ are homotopic if and only if they have the same degree, $\deg f = \deg g$.

3. Prove that if we glue a Möbius strip to a two-dimensional disk along the identity diffeomorphism of their boundaries, then we obtain the projective plane $\mathbb{R}P^2$.

4. Prove that the connected sum of the torus and the projective plane is isomorphic to the connected sum of three projective planes:

$$T^2\#\mathbb{R}P^2 = \mathbb{R}P^2\#\mathbb{R}P^2\#\mathbb{R}P^2.$$

5. Consider the square $0 \le x, y \le 1$ on the plane and identify the boundary intervals by the rule

$$(0,y) \sim (1, 1-y), \quad (x,0) \sim (x,1).$$

The surface thus obtained is called the *Klein bottle.* Prove that it is homeomorphic to the connected sum of two projective planes $\mathbb{R}P^2\#\mathbb{R}P^2$.

6. Prove that if $L$ contains higher order derivatives,

$$L = L(t, x, \dot{x}, \ddot{x}, \dots, x^{(k)}),$$

then the Euler–Lagrange equation for the functional $S[\gamma] = \int_\gamma L\,dt$ has the form

$$\frac{\delta S}{\delta x} = \frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} + \frac{d^2}{dt^2}\frac{\partial L}{\partial \ddot{x}} + \cdots + (-1)^k \frac{d^k}{dt^k}\frac{\partial L}{\partial x^{(k)}} = 0.$$

7. Prove that the trajectories of a particle moving in a central force field $U = U(r)$ in $\mathbb{R}^3$ are planar curves.

*Chapter 13*

# Poisson and Lagrange Manifolds

## 13.1. Symplectic and Poisson manifolds

**13.1.1. $g$-gradient systems and symplectic manifolds.** We begin with local theory, having in mind a large class of manifolds. In any domain of the space $\mathbb{R}^m$ with coordinates $(y^1, \dots, y^m)$ the most important geometric structures can be specified by means of a scalar product of vectors or covectors $g_{ij}$ or $g^{ij}$. We do not impose any symmetry on them at the moment. In the nondegenerate case it is assumed that

$$g_{jk}g^{ki} = g^{ik}g_{kj} = \delta^i_j,$$

i.e., these matrices are mutually inverse. To begin with, suppose that this condition if fulfilled.

In this structure, the gradient $\nabla f$ of a function $f(y^1, \dots, y^m)$ is defined as the vector

$$(\nabla f)^i = g^{ij}\,\frac{\partial f}{\partial y^j}, \tag{13.1}$$

without assuming that the tensor $g^{ij}$ is symmetric. To the vector field $\nabla f$ there corresponds the gradient dynamic system, i.e., the system of equations (see 8.3.1)

$$\dot{y}^i = (\nabla f)^i. \tag{13.2}$$

A system of the form (13.2) is called a *$g$-gradient system*, with the function $f$ being its *generator*. In general, generators of Hamiltonian systems are also called *Hamiltonians*. We have the following simple lemma.

**Lemma 13.1.** *For any function $h(y)$, its derivative $\nabla f(h)$ has the form*

$$\dot{h} = \nabla f(h) = \langle \nabla h, \nabla f \rangle = g^{ij} \frac{\partial h}{\partial y^i} \frac{\partial f}{\partial y^j}.$$

Here and subsequently we denote by $\nabla f(h)$ the directional derivative of $h$ in the direction of the field $\nabla f$ of the form (13.1).

**Proof.** By definition,

$$\dot{h} = \frac{\partial h}{\partial y^i} \dot{y}^i = \frac{\partial h}{\partial y^i} (\nabla f)^i = \frac{\partial h}{\partial y^i} g^{ij} \frac{\partial f}{\partial y^j}.$$

Hence the lemma. □

Of particular interest for us will be nondegenerate skew-symmetric "metrics" $g_{ij} = -g_{ji}$ specified by 2-forms

$$\Omega = \sum_{i<j} g_{ij}\, dy^i \wedge dy^j,$$

for which the inverse matrix $g^{ij}$ is defined, so that $g_{ij} g^{jk} = \delta_i^k$, $g^{ij} = -g^{ji}$, and $\det g_{ij} = g \neq 0$. Obviously, the space in this case has an even dimension, $i, j = 1, \ldots, 2n$ (see 2.1.1).

This interest has fundamental grounds: the generator of any such system is a conservation law, since

$$\frac{df}{dt} = \langle \nabla f, \nabla f \rangle = 0$$

by the skew-symmetry condition $g^{ij} = -g^{ji}$ (*energy conservation law*).

**Lemma 13.2.** *The following formula holds*:

$$\frac{1}{n!} \underbrace{\Omega \wedge \cdots \wedge \Omega}_{n} = \sqrt{g}\, dy^1 \wedge \cdots \wedge dy^{2n}. \tag{13.3}$$

*In particular, $\sqrt{g}$ is a polynomial function of $g_{ij}$ (called a* Pfaffian*).*

**Proof.** It suffices to prove (13.3) for each separate point. Hence without loss of generality we may assume that $g_{ij} = \text{const}$. Introduce new coordinates

$$(x^1, \ldots, x^n, p_1, \ldots, p_n) = (z^1, \ldots, z^{2n})$$

in which $g_{ij} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ (here 1 is the identity $n \times n$-matrix). In this case

$$\Omega = \sum_i dp_i \wedge dx^i,$$

$$\Omega \wedge \cdots \wedge \Omega = n!\, dz^1 \wedge \cdots \wedge dz^{2n}.$$

In these coordinates the lemma is proved, since $\sqrt{g} = 1$. In view of invariance of formula (13.3), the lemma holds in any coordinate system. □

Thus the nondegeneracy condition $g \neq 0$ is equivalent to the condition $\Omega^n \neq 0$.

A manifold is said to be endowed with a *symplectic structure* if on this manifold a nondegenerate 2-form (a nondegenerate skew-symmetric scalar product $g_{ij}$ of tangent vectors at each point $x$) is specified, such that for any point $x$ there exists its neighborhood $U$ with local coordinates $(y^1, \dots, y^n, y^{n+1}, \dots, y^{2n})$ in which the scalar product can be written as

$$g_{ij} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad g^{ij} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

or the form is given by the "canonical" expression

$$\sum_{i<j} g_{ij}\, dy^i \wedge dy^j = \Omega = \sum dp_\alpha \wedge dx^\alpha, \tag{13.4}$$

where $p_\alpha = y_{n+\alpha}, x^\alpha = y^\alpha, \alpha = 1, \dots, n$.

The form $\Omega$ is called a *symplectic structure* or a *symplectic form* on the manifold.

The next lemma, obviously, follows directly from the definition.

**Lemma 13.3.** *If a manifold $M^{2n}$ is endowed with a symplectic structure $\Omega$, then the form $\Omega$ is closed, $d\Omega = 0$.*

The converse statement is also true.

**Theorem 13.1** (Darboux). *Any closed nondegenerate 2-form can be reduced locally to the canonical form* (13.4), *and so any closed nondegenerate 2-form specifies a symplectic structure.*

We postpone the proof of this theorem.

A symplectic manifold, i.e., a space with a skew-symmetric metric $g_{ij}$, which admits, in a neighborhood of any of its points, local coordinates $(x, p)$ such that

$$g_{ij} \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

is also called a *phase space.*

The coordinates $(x, p)$ in which the form $\Omega = \sum_{i<j} g_{ij}\, dy^i \wedge dy^j$ is written as in (13.4) are called *canonical.*

$g$-gradient systems on symplectic manifolds are called *Hamiltonian systems.*

Hamiltonian systems have the form

$$\dot{y}^i = (\nabla H)^i, \quad i = 1, \dots, 2n,$$

or, in canonical coordinates,

$$\dot{x}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial x^i}, \quad i = 1, \dots, n. \tag{13.5}$$

Lemma 13.1 entails the following lemma.

**Lemma 13.4.** *For a Hamiltonian system* (13.5) *the derivative of any function* $f(x, p, t)$ *equals*

$$\dot{f} = \frac{\partial f}{\partial t} + \langle \nabla f, \nabla H \rangle,$$

*where, in canonical coordinates,*

$$\langle \nabla f, \nabla H \rangle = g^{ij} \frac{\partial f}{\partial y^i} \frac{\partial H}{\partial y^j} = \sum_{i=1}^{n} \Big( \frac{\partial f}{\partial x^i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial x^i} \frac{\partial f}{\partial p_i} \Big).$$

*In particular, if* $f = H = H(x, p, t)$, *then*

$$\dot{E} = \dot{H} = \frac{\partial H}{\partial t},$$

*since* $\langle \nabla H, \nabla H \rangle = -\langle \nabla H, \nabla H \rangle = 0$.

A function $f(x, p)$ on the phase space (which does not explicitly depend on time) is said to be a *first integral* or simply an *integral* (*of motion*) *of the Hamiltonian system* if it is preserved along the trajectories of the system:

$$\dot{f} = 0.$$

Since the system is Hamiltonian, this is equivalent to the identity

$$\dot{f} = \langle \nabla f, \nabla H \rangle = 0.$$

If the Hamiltonian $H$ of the system does not explicitly depend on time, then, as the preceding lemma shows, it is an integral of this system.

The importance of the class of Hamiltonian systems is due to the theorem on equivalence of the Euler–Lagrange equations (for strongly nondegenerate Lagrangians) and the Hamilton equations in the phase space (see 12.4.1).

**13.1.2. Examples of phase spaces.** The most common example of a phase space in modern mechanics and physics is the space $T^*M$ of all covectors on some $n$-dimensional manifold $M$. Its points $y \in T^*M$ are pairs

$$y = (x, \xi),$$

where $x \in M$ and $\xi$ is a covector at the point $x$. The dimension of this space is $2n$. Local coordinates $(y^1, \dots, y^{2n})$ about a point $y \in T_*M$ are specified as $(x^1, \dots, x^n, \xi_1, \dots, \xi_n)$, where $x^i$ are local coordinates on $M$ about the point $x \in M$, and $(\xi_1, \dots, \xi_n)$ are the coordinates of the covector $\xi$ in this system.

These local coordinates $(x, \xi)$ on $T^*M$ are *canonical*: the 2-form

$$\Omega = \sum_{\alpha=1}^{n} d\xi_\alpha \wedge dx^\alpha$$

is invariant (i.e., does not depend on local coordinates $x^1, \dots, x^n$). It specifies the *standard symplectic structure* on $T^*M$. The manifold $M$ is the "configuration space".

EXAMPLES. 1. For the problem of $m$ interacting points $x_i$ in $\mathbb{R}^3$, $i = 1, \dots, m$, we have $M = \mathbb{R}^{3m}$ and $T^*M = \mathbb{R}^{6m}$ with coordinates

$$(x_1^1, x_1^2, x_1^3, \dots, x_m^1, x_m^2, x_m^3, p_{11}, p_{12}, p_{13}, \dots, p_{m1}, p_{m2}, p_{m3})$$

and the standard 2-form $\Omega = \sum dp_\alpha \wedge dx^\alpha$.

2. For the motion of a solid body about a fixed point, we have

$$M = \mathrm{SO}(3), \quad T^*M = \mathrm{SO}(3) \times \mathbb{R}^3,$$

since the position of the solid body is uniquely determined by some orthonormal frame fixed to the body.

3. For the motion of a solid body in space, we have

$$M = \mathrm{SO}(3) \times \mathbb{R}^3, \quad T^*M = \mathrm{SO}(3) \times \mathbb{R}^3 \times \mathbb{R}^6.$$

Later on, we will say more about the natural Lagrangians describing the motion of a solid body. In the case of free motion of a body in space, it can be easily shown that the mass center of the body moves at a constant velocity. In the coordinate system moving together with the body (i.e., when the mass center is fixed) this reduces to the previous case, where the fixed point lies in the mass center (Euler's motion). Consider this case of a "free gyroscope". The motion is described by a curve in the group SO(3):

$$g(t) \in \mathrm{SO}(3).$$

On the Lie algebra of skew-symmetric matrices the *tensor of inertia* is specified, which is a positive quadratic form $\langle , \rangle_I$ determined by the distribution of mass in the body. The Lagrangian has the form

$$L(g, \dot{g}) = \langle A(t), A(t) \rangle_I,$$

where $A(t)$ is the "angular velocity",

$$A(t) = g^{-1}(t) \, \frac{dg(t)}{dt},$$

regarded as a skew-symmetric matrix (an element of the Lie algebra of the group SO(3)).

We see that this Lagrangian possesses the symmetry

$$L(g_0 g, g_0 \dot{g}) = L(g, \dot{g})$$

(left-invariance). Indeed,

$$(g_0g)^{-1}\,\frac{d}{dt}\,(g_0g) = g^{-1}g_0^{-1}g_0\dot{g} = g^{-1}\dot{g} = A.$$

Therefore, $\langle A, A\rangle_I$ stays unchanged.

Thus we arrive at the following conclusion.

**Lemma 13.5.** *Any left-invariant Lagrangian $L(g_0g, g_0\dot{g}) = L(g, \dot{g})$ on the group $G = \mathrm{SO}(3)$ is determined by a single real-valued function $\varepsilon(A)$ on the Lie algebra, where $A = g^{-1}\dot{g}$ is a skew-symmetric matrix.*

Actually, we have already proved this for $L = \langle A, A\rangle_I$.

Now we present an important *nonstandard symplectic structure on $T^*M$ constructed for a magnetic field* on the manifold $M$.

Consider a charged particle with charge $e$ in $\mathbb{R}^3$, with Lagrangian $L_0(x, \dot{x})$ and Hamiltonian $H_0(x, p)$, when there is no magnetic field. After inclusion of a stationary magnetic field specified by a 2-form

$$B = \sum_{\alpha<\beta} B_{\alpha\beta}\, dx^\alpha \wedge dx^\beta, \quad \alpha, \beta = 1, 2, 3,$$

we can locally introduce the vector-potential (independent of time)

$$dA = B$$

and write down the Lagrangian and the action as

$$L = L_0(x, \dot{x}) + \frac{e}{c}\, A_\alpha(x)\dot{x}^\alpha,$$

$$S = \int L_0(x, \dot{x}) + \frac{e}{c}\int A_\alpha(x)\, dx^\alpha.$$

However, this formula is of local nature. What is the global meaning of this? We will return to this question later, but one of the answers is as follows.

**Theorem 13.2.** *Introduce the new symplectic form*

$$\Omega = \sum_\alpha dp_\alpha \wedge dx^\alpha + \frac{e}{c}\, B$$

*on the space $T^*\mathbb{R}^3$. Then the same Hamiltonian $H_0(x, p)$ as before in the new symplectic structure specifies the equation of motion of the charged particle in the magnetic field.*

REMARK. As is usually stated in physics textbooks, inclusion of the magnetic field does not affect energy. This principle in the present geometric form has apparently been formulated for the first time by one of the authors of this book.

**Proof.** The proof consists in elementary verification.

In the absence of magnetic field we have

$$\frac{d}{dt}\Big(\frac{\partial L_0}{\partial \dot{x}^\alpha}\Big) = \frac{\partial L_0}{\partial x^\alpha}.$$

After inclusion of the magnetic field we obtain

$$\frac{d}{dt}\Big(\frac{\partial L_0}{\partial \dot{x}^\alpha}\Big) = \frac{\partial L_0}{\partial x^\alpha} + \frac{e}{c}\Big[\frac{\partial A_\beta}{\partial x^\alpha}\dot{x}^\beta - \frac{\partial A_\alpha}{\partial x^\beta}\dot{x}^\beta\Big] = \frac{\partial L_0}{\partial x^\alpha} + \frac{e}{c} B_{\alpha\beta}\dot{x}^\beta,$$

where $B_{\alpha\beta} = \frac{\partial A_\beta}{\partial x^\alpha} - \frac{\partial A_\alpha}{\partial x^\beta}$.

The inverse matrix $g^{ij}$ to the new symplectic form $\Omega$ is

$$g^{ij} = \begin{pmatrix} 0 & 1 \\ -1 & \frac{e}{c}B_{\alpha\beta}(x) \end{pmatrix},$$

where $i, j = 1, 2, \ldots, 6$, $\alpha, \beta = 1, 2, 3$. Now we find the $g$-gradient of the Hamiltonian $H_0(x, p)$ (see (13.5)). As before, we have

$$\dot{x}^\alpha = \frac{\partial H_0}{\partial p_\alpha} = g^{\alpha j}\frac{\partial H_0}{\partial y^j}, \quad \alpha = 1, 2, 3,$$
$$(y^1, y^2, y^3, y^4, y^5, y^6) = (x^1, x^2, x^3, p_1, p_2, p_3).$$

For the second part of equations (13.5) we obtain

$$\dot{p}_\alpha = -\frac{\partial H_0}{\partial x^\alpha} + \frac{e}{c}B_{\alpha\beta}\frac{\partial H_0}{\partial p^\beta}.$$

Combined with the first part, this provides the equation of the motion in the magnetic field. The proof is completed. ☐

From this point of view, the introduction (locally) of the vector-potential

$$B = d(A_\alpha(x)\, dx^\alpha) = \sum_{\alpha<\beta} B_{\alpha\beta}\, dx^\alpha \wedge dx^\beta$$

is simply introduction of new canonical coordinates.

**Lemma 13.6.** *Let*

$$\tilde{p}_\alpha = p_\alpha + \frac{e}{c}A_\alpha,$$

*where $d(A_\alpha\, dx^\alpha) = B$. Then the form $\Omega$ coincides with the standard form*

$$\Omega_0 = \sum d\tilde{p}_\alpha \wedge dx^\alpha.$$

**Proof.** By definition, we have

$$\sum d\tilde{p}_\alpha \wedge dx^\alpha = \sum (dp_\alpha + \frac{e}{c}\, dA_\alpha) \wedge dx^\alpha = \sum dp_\alpha \wedge dx^\alpha + \frac{e}{c}\, dA_\alpha \wedge dx^\alpha.$$

At the same time,

$$d(A_\alpha\, dx^\alpha) = \sum_{\alpha<\beta} B_{\alpha\beta}\, dx^\alpha \wedge dx^\beta.$$

Hence the lemma. □

If $[\Omega] = \frac{e}{c}[B]$ is a nonzero cohomology class, then the global reduction to the canonical form is impossible. However, expressions of the form $d^{-1}(\Omega)$ are locally well defined if the form $\Omega$ is closed. This follows from the Poincaré lemma (see 9.3.2).

Expressions of the form $d^{-1}(\Omega)$ arise in the formulation of the following *global variational Hamilton's principle* on an arbitrary symplectic manifold (phase space), i.e., on a manifold $M$ with a symplectic form $\Omega$ which is locally reducible to the form $\Omega = \sum dp_\alpha \wedge dx^\alpha$ in special local canonical coordinates $y = (x, p)$ in a neighborhood of any point $(y) \in M$.

**Lemma 13.7.** *Choose locally a* 1*-form* $\omega$ *such that* $d\omega = \Omega$. *Then the expression* (*integral along the curve*)

$$S = \int (\omega - H\, dt) = \int (d^{-1}(\Omega) - H\, dt)$$

*has an extremum on any curve where the Hamilton equations are fulfilled* (*i.e., the curve is a trajectory of a Hamiltonian system*) *with Hamiltonian* $H$ *and symplectic form* $\Omega$. *In particular, this holds for all strongly local variations of the curve concentrated in a small neighborhood of the point under consideration where the* 1*-form* $\omega$ *is defined.*

**Proof.** The Euler–Lagrange equations remain unchanged if we make a gradient change $\omega \to \omega + d\varphi$, i.e., add a total derivative. Hence we may assume without loss of generality that the 1-form $\omega$ is locally chosen as $\sum_\alpha p_\alpha\, dx^\alpha = \omega$. For this choice of the 1-form the lemma has already been proved before as Hamilton's variational principle (Theorem 12.22). This implies also the proof in the general case. □

Later on, we will return to the question about the meaning of the invariant expression

$$S = \int (d^{-1}(\Omega) - H\, dt)$$

on a symplectic manifold. It is worthwhile to note that, as the above example shows, expressions of the form $d^{-1}(\Omega)$ arise in real physical applications.

Now we introduce the notion of a *closed* 1*-form on an arbitrary topological space* $X$. Suppose we are given an arbitrary covering of the space $X$ by open sets,

$$X = \bigcup_q V_q.$$

A collection of continuous functions $\varphi_q\colon V_q \to \mathbb{R}$ will be referred to as the *antiderivative* of a closed 1-form on $X$ that is exact in each domain $V_q$, if in the intersections $V_q \cap V_s$ the differences $\varphi_q - \varphi_s$ are locally constant functions (constant on each connected component).

Consider some examples.

1. Let $N = X$ be a manifold and $\omega$ a 1-form on it. Consider a covering

$$N = \bigcup_q V_q$$

fine enough so that on each domain $V_q$,

$$\omega = d\varphi_q(x), \quad x \in V_q.$$

For example, if the domains $V_q$ are disk-shaped convex domains in $\mathbb{R}^n$, then we simply take one point $x_q$ in each $V_q$ and set

$$\varphi_q(x) = \int_{x_q}^{x} \omega,$$

where the integration path lies entirely in $V_q$. Then for the differences in the intersections $V_q \cap V_s$ we obtain

$$d[\varphi_q(x) - \varphi_s(x)] = 0.$$

Hence these differences are locally constant.

2. Consider now the space $X$ whose points are all smooth mappings of the circle $S^1$ with coordinate $t \pmod{2\pi}$ into the manifold $M$ under consideration with a closed 2-form $\Omega$, $d\Omega = 0$. This is the *space of loops* $L(M) = X$.

**Lemma 13.8.** *For any smooth curve $\gamma' \in L(M) = X$ there is an open domain $V_{\gamma'}$ in the space of curves $X$ such that the expression*

$$\int_\gamma d^{-1}(\Omega) = \varphi_{\gamma'}(\gamma)$$

*is well defined as a functional on the domain $V_{\gamma'}$ of the space of curves $X$.*

**Proof.** By the transversality theorems, the image of the circle $S^1$ under an infinitely differentiable mapping $\gamma'\colon S^1 \to M$ covers a one-dimensional set $\gamma(S^1) \subset M$. We will use the following fact without proof: for any one-dimensional compact closed set in a smooth manifold $M$, there is a sufficiently small open neighborhood $U_{\gamma'} \supset \gamma(S^1)$ such that the restriction of the 2-form $\Omega$ to it is exact:

$$\Omega = d\omega_{\gamma'} \quad (\text{in the domain } U_{\gamma'}).$$

This is obviously true for a polygonal mapping of the circle into a domain of Euclidean space. Denote by $V_{\gamma'} \subset X$ the open domain in the space

$X = L(M)$ consisting of all smooth mappings $\gamma\colon S^1 \to U_{\gamma'}$. For any curve $\gamma \in V_{\gamma'}$ we set by definition

$$S_{\gamma'}(\gamma) = \oint_\gamma \omega_{\gamma'}.$$

The proof (though somewhat informal) is completed. □

The domains $V_{\gamma'}$ cover the space $X$, since the curve $\gamma'$ may be chosen arbitrarily. They form a set of continuum cardinality.

**Theorem 13.3.** *The expression $\int d^{-1}(\Omega)$ determines a well-defined closed 1-form on the space $X = L(M)$ of smooth closed curves in the manifold $M$.*

**Proof.** Using the lemma we construct a continual covering

$$X = \bigcup_q V_q,$$

where $q = \gamma' \in X = L(M)$, i.e., for each curve $\gamma'$ we construct a covering domain with the required property.

Consider the intersection $\gamma \in V_q \cap V_s$, where $q = \gamma_1$, $s = \gamma_2$ are two curves. Then there exist two forms $\omega_{\gamma_1} = \omega_1$ and $\omega_{\gamma_2} = \omega_2$ such that $d\omega_1 = d\omega_2 = \Omega$.

Consider the difference $\varphi_1 - \varphi_2$:

$$\varphi_1 - \varphi_2 = \int_{\gamma'} \omega_1 - \int_{\gamma'} \omega_2 = \int_{\gamma'} (\omega_1 - \omega_2).$$

Since the form $\omega_1 - \omega_2$ is closed, the integral over a closed path

$$\oint (\omega_1 - \omega_2)$$

remains unchanged under a small deformation of this path. Therefore, the functional (function) $\varphi_1 - \varphi_2$ on the intersection $V_{\gamma_1} \cap V_{\gamma_2}$ of the domains in the space of curves is locally constant. Hence the theorem. □

REMARK. We see that the antiderivative of a closed form on the space of curves is well defined by the expression $\int d^{-1}(\Omega)$. In a naive way, this can be stated as follows: as we know, the variation $\delta S$ of a functional is well defined. It is a closed form. It determines the Euler–Lagrange and Hamilton equations, but the functional itself is not, in general, defined globally. We will deal with topological consequences of this phenomenon later on.

**13.1.3. Extended phase space.** Let $H = H(x,p,t)$ be a Hamiltonian that explicitly depends on time $t$. It determines a Hamiltonian system on a phase space $M$ with local coordinates $x^i, p_j$, $i,j = 1,2,\dots,n$.

From the Hamilton equations

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x},$$

we obtain the equations

$$\dot{E} = \frac{\partial H}{\partial t}, \quad \dot{t} = 1.$$

It is expedient to introduce the so-called *extended phase space* with local coordinates $(x,p,E,t)$, where $x^{n+1} = E$, $p_{n+1} = t$, and with skew-symmetric metric

$$\hat{g}_{ij} = \begin{pmatrix} 0 & -1 & & & & & 0 \\ 1 & 0 & & & & & \\ & & \ddots & & & & \\ & & & 0 & -1 & & \\ & & & 1 & 0 & & \\ & & & & & 0 & -1 \\ 0 & & & & & 1 & 0 \end{pmatrix},$$

to which corresponds the form

$$\hat{\Omega} = \sum_{i=1}^{n} dp_i \wedge dx^i - dE \wedge dt.$$

Consider the Hamiltonian $\hat{H}(x,p,y,E) = H(x,p,t) - E$ and the corresponding Hamiltonian system

$$\dot{x} = \frac{\partial \hat{H}}{\partial p}, \quad \dot{p} = -\frac{\partial \hat{H}}{\partial x}, \quad \dot{t} = -\frac{\partial \hat{H}}{\partial E}, \quad \dot{E} = \frac{\partial \hat{H}}{\partial t}.$$

This implies the following theorem.

**Theorem 13.4.** *In the extended phase space $(x,p,t,E)$ with metric $\hat{g}_{ij}$ (or form $\hat{\Omega}$) the Hamilton equations with Hamiltonian $\hat{H} = H(x,p,t) - E$ on the surface $\hat{H} = 0$ coincide with the initial Hamilton equations in the space $(x,p)$ supplemented with the relations $\dot{t} = 1$, $\dot{E} = \partial H/\partial t$.*

Now the energy conservation law is also explained by symmetry. If the initial Hamiltonian $H$ does not explicitly depend on time, then the time $t = x^{n+1}$ is a cyclic coordinate and the corresponding momentum $p_{n+1} = E$ is conserved.

**13.1.4. Poisson manifolds and Poisson algebras.** To construct Hamiltonian systems, we need only a skew-symmetric scalar product $g^{ij} = -g^{ji}$ with superscripts. If a manifold $M$ is endowed with a skew-symmetric tensor field $g^{ij} = -g^{ji}$, then the $g$-gradient system

$$\dot{x}^i = g^{ij} \frac{\partial H}{\partial x^j}$$

is called a *Hamiltonian system*, and the function $H$ is called the *Hamiltonian*. In this setting the scalar product of covectors $(g^{ij})$ may be degenerate: we only assume that $\det(g^{ij}) \neq 0$, admitting the possibility of $\det(g^{ij}) = 0$.

A pair $(M, g^{ij})$ consisting of a manifold $M$ and a skew-symmetric tensor field $(g^{ij} = -g^{ji})$ is called a *Poisson manifold* or a *Poisson structure on the manifold* $M$ if for any pair of smooth real functions $f, h$ on $M$, their *Poisson bracket* ($g$-scalar product of their gradients) is defined,

$$\{f, g\} = g^{ij}(x) \frac{\partial f}{\partial y^i} \frac{\partial g}{\partial y^j},$$

where $(y^1, \dots, y^m)$ are local coordinates and any triple of real smooth functions $f, g, h$ on $M$ satisfies the Jacobi identity

$$0 = \{\{f, g\}, h\} + \{\{h, f\}, g\} + \{\{g, h\}, f\}.$$

The Poisson brackets possess the following important properties:

$$\{f, g\} = -\{g, f\} \quad \text{(skew-symmetry)},$$
$$\{fg, h\} = f\{g, h\} + g\{f, h\} \quad \text{(Leibniz identity)}.$$

Now we introduce the general concept of the Poisson algebra over an arbitrary field.

A linear space $C$ over an arbitrary field is called a *Poisson algebra* if it is equipped with two additional operations:

a) multiplication of elements $ab$, $a \in C$, $b \in C$, which makes $C$ a commutative associative algebra over this field, with the identity $1 \in C$;

b) a bilinear "Poisson bracket" $\{a, b\}$, which is skew-symmetric,

$$\{a, b\} = -\{b, a\},$$

related to the above multiplication by the Leibniz identity,

$$\{ab, c\} = a\{b, c\} + b\{a, c\},$$

and satisfies the Jacobi identity for any triple of elements,

$$\sum_{\text{cyclic}} \{\{a, b\}, c\} = 0$$

(where the summation is over all cyclic permutations of the elements $a$, $b$, and $c$).

First we consider *pre-Poisson* algebras $C$, which have all the properties except for the Jacobi identity.

Moreover, we will assume that the algebra $C$ is generated over the field $K$ by a set of elements $x^1, x^2, \ldots \in C$, i.e., any element $x \in C$ is expressed as a polynomial in $x^1, x^2, \ldots$.

**Lemma 13.9.** *For any polynomial in $n$ variables $\Phi(a_1, \ldots, a_n)$, $a_j \in C$, and any element $b \in C$, the general Leibniz identity holds:*

$$\{\Phi(a_1, \ldots, a_n), b\} = \sum_j \frac{\partial \Phi}{\partial a_j} \{a_j, b\}.$$

**Proof.** This fact for arbitrary polynomials $\Phi(a_1, \ldots, a_n)$ is proved similarly to the usual derivation of this identity for partial derivatives

$$\{a_j, b\} \leftrightarrow \frac{\partial}{\partial x^j}$$

using the property

$$\frac{\partial}{\partial x^j}(fg) = \frac{\partial f}{\partial x^j} g + f \frac{\partial g}{\partial x^j}$$

from elementary calculus. The operators $\{a_j, b\}$ possess the same properties as $\frac{\partial}{\partial x^j}$. Hence the lemma. □

The following lemma states a condition for an algebra $C$ to be a Poisson algebra.

**Lemma 13.10.** *If the Jacobi identity is fulfilled for any triple of generators,*

$$\sum_{\text{cyclic}} \{\{x^i, x^j\}, x^k\} = 0,$$

*then it holds for any triple of elements $f, g, h \in C$ that are polynomials in $x^1, x^2, \ldots$.*

**Proof.** Lemma 13.9 implies that

$$\sum_{\text{cyclic}} \{\{f, g\}, h\} = \frac{\partial f}{\partial x^i} \frac{\partial g}{\partial x^j} \frac{\partial h}{\partial x^k} \Big( \sum_{\text{cyclic}} \{\{x^i, x^j\}, x^k\} \Big)$$

(see the proof of Lemma 6.11 in 6.1.9). Hence the lemma. □

Now we proceed to algebras of functions in which the Poisson bracket operation is specified by a smooth tensor field $(g^{ij})$ on a manifold $M^n$. Similarly to usual differentiation, smooth analogs of Lemmas 13.9 and 13.10 are valid. We state without proof that Lemma 13.9 holds for any smooth functions of coordinates:

$$\{f, g\} = \frac{\partial f}{\partial x^i} \{x^i, g\}.$$

We have already introduced Poisson manifolds in 6.1.9 when discussing examples of infinite-dimensional Lie algebras, without any connection with the Hamiltonian formalism. Therefore, we will sometimes recall some facts with references to the previous proofs.

Instead of Lemma 13.10 we have the following theorem (it was proved in 6.1.9; see Lemma 6.11).

**Theorem 13.5.** *A Poisson structure is well defined by a tensor field* $(g^{ij})$ *on a manifold* $M$ *if and only if the following identity holds:*

$$\sum_{\substack{\text{cyclic}\\(ijk)}} \frac{\partial g^{ij}}{\partial y^l}\, g^{lk} = 0.$$

**Corollary 13.1.** *Let the Poisson tensor* $g^{ij}$ *be a constant matrix in coordinates* $(y)$. *Then the Poisson brackets*

$$\{f, g\} = g^{ij}(y)\, \frac{\partial f}{\partial y^i}\, \frac{\partial g}{\partial y^j}$$

*satisfy the Jacobi identity for any two functions* $f, g \in C^\infty(M, \mathbb{R})$.

EXAMPLE. Let a symplectic structure be specified in canonical coordinates $(x^\alpha, p_\alpha)$, $\alpha = 1, \ldots, n$, $\Omega = \sum dp_\alpha \wedge dx^\alpha$, by the formula

$$g^{ij} = \begin{pmatrix} 0 & 1_n \\ -1_n & 0 \end{pmatrix}.$$

Then the Poisson bracket becomes

$$\{f(x,p), g(x,p)\} = \sum_{\alpha=1}^{n} \Big( \frac{\partial f}{\partial x^\alpha}\, \frac{\partial g}{\partial p_\alpha} - \frac{\partial g}{\partial x^\alpha}\, \frac{\partial f}{\partial p_\alpha} \Big).$$

It satisfies the Jacobi identity.

**Theorem 13.6.** *A Poisson structure on a manifold* $M^{2n}$, $i, j = 1, \ldots, 2n$, *is well defined by a nondegenerate skew-symmetric scalar product* $g^{ij}$, $\det(g^{ij}) \neq 0$, *if and only if the corresponding nondegenerate form* $\Omega = \sum_{i<j} g_{ij}\, dy^i \wedge dy^j$ *is closed,*

$$d\Omega = 0.$$

**Proof.** We proved in 6.1.9 that a nondegenerate skew-symmetric scalar product $g^{ij}$ defines a Poisson structure if and only if

$$\frac{\partial g_{rs}}{\partial y^p} + \frac{\partial g_{sp}}{\partial y^r} + \frac{\partial g_{pr}}{\partial y^s} = 0, \quad p, r, s = 1, \ldots, n,$$

where index raising and lowering are done by means of a skew-symmetric metric, $g^{ij} g_{jk} = \delta^i_k$.

Obviously, these relations are equivalent to the equation

$$d\Omega = 0$$

for the 2-form $\Omega = \sum_{i<j} g_{ij}\, dy^i \wedge dy^j$. Hence the theorem. □

Recall an important operation: the commutator of vector fields on a manifold $M$ (see 8.3.2).

To a vector field $\eta$ corresponds a first-order operator acting on functions (the directional derivative):

$$\partial_\eta f = \eta^i \frac{\partial f}{\partial y^i}$$

in local coordinates $(y^1, \dots, y^m)$. The commutator of two operators is also a first-order operator:

$$\partial_\eta \partial_\xi - \partial_\xi \partial_\eta = \partial_{[\eta,\xi]}.$$

The vector field $[\eta, \xi]$ is called the *commutator of the fields* $\eta$ and $\xi$ on $M$. The field $[\eta, \xi]$ has the form

$$[\eta, \xi]^i = \eta^j \frac{\partial \xi^i}{\partial y^j} - \xi^j \frac{\partial \eta^i}{\partial y^j}.$$

**Theorem 13.7.** *The Poisson brackets possess the following property*:

$$\nabla\{f, g\} = -[\nabla f, \nabla g],$$

*where* $[\,,]$ *is the commutator of vector fields and* $(\nabla h)^i = g^{ij} \frac{\partial h}{\partial x^j}$ *is the Hamiltonian vector field corresponding to the Hamiltonian* $h$.

**Proof.** By the definition of the Hamiltonian vector field and Poisson brackets we have

$$\nabla f(g) = \{g, f\} = -\{f, g\}.$$

Hence

$$\nabla\{f, g\}(h) = -\{\{f, g\}, h\}.$$

The Jacobi identity for this triple implies that

$$\nabla\{f, g\}(h) - \{\{h, f\}, g\} - \{\{g, h\}, f\} = 0.$$

Using skew-symmetry, we obtain

$$\begin{aligned}\nabla\{f, g\}(h) &= \nabla g(\{h, f\}) - \nabla f(\{h, g\} \\ &= \nabla g \nabla f(h) - \nabla f \nabla g(h) = [\nabla g, \nabla f](h).\end{aligned}$$

The proof is completed. □

Now we introduce an important notion. The *center* $\mathbb{Z}_C$ of a Poisson algebra $C$ is the set of elements $z$ such that

$$\{z, C\} = 0.$$

The center $\mathbb{Z}_C$ is also called the *annihilator* of the Poisson algebra, and its elements $z \in \mathbb{Z}_C$ are called *Casimir functions.*

Suppose in some phase space we have a Hamiltonian system with Hamiltonian $H$. Then the derivative of any function $f = f(x,p)$ along this system has the form

$$\dot{f} = \{f, H\}. \tag{13.6}$$

In particular,

$$\dot{x}^i = \{x^i, H\}, \quad \dot{p}_i = \{p_i, H\}.$$

Therefore, a function $f(x,p)$ is an integral of the Hamiltonian system (integral of motion) if it commutes with the Hamiltonian $H$:

$$\{f, H\} = 0.$$

The set of integrals of a Hamiltonian system is a Lie algebra, which is also closed relative to multiplication of functions. Hence we obtain the following lemma.

**Lemma 13.11.** *The integrals of motion of a Hamiltonian system form a Poisson algebra.*

The algebra of integrals of motion is also denoted by $Z_H$ and called the *centralizer* of $H$. It follows from the Jacobi identity that the centralizer $Z_H$ itself is a Poisson algebra because the Poisson bracket of integrals of motion is also an integral of motion. Casimir functions are integrals for any Hamiltonian system (with any Hamiltonian).

EXAMPLE. Let $L(x, \dot{x})$ be a spherically symmetric Lagrangian. As was shown in 12.3.4, to the three one-parameter subgroups of SO(3) correspond three moment integrals $M_x$, $M_y$, $M_z$, where

$$M_x = L_x^i \frac{\partial L}{\partial \dot{x}^i}, \quad M_y = L_y^i \frac{\partial L}{\partial \dot{x}^i}, \quad M_z = L_z^i \frac{\partial L}{\partial \dot{x}^i},$$

with $L_x$, $L_y$, $L_z$ being the corresponding linear vector fields. The commutators of these fields are (see 8.3.3)

$$[L_x, L_y] = L_z, \quad [L_y, L_z] = L_x, \quad [L_z, L_x] = L_y.$$

Theorem 13.7 implies that the Poisson brackets of the components of a moment are given by the formulas:

$$\{M_x, M_y\} = M_z, \quad \{M_y, M_z\} = M_x, \quad \{M_z, M_x\} = M_y.$$

Therefore, the functions $M_x$, $M_y$, and $M_z$ on the phase space of a spherically symmetric system form a Lie algebra relative to Poisson brackets, which is

isomorphic to the Lie algebra $\mathfrak{so}(3)$. Polynomials in these functions form a nontrivial Poisson algebra.

**13.1.5. Reduction of Poisson algebras.** Let $H$ be a Hamiltonian on a Poisson manifold $M$, and suppose that $H$ does not explicitly depend on time, $\frac{\partial H}{\partial t} = 0$. Then the energy is conserved and all the trajectories of the Hamiltonian system with Hamiltonian $H$ lie on a level surface $H = E$.

Let $f$ be an integral of this system, $\{f, H\} = 0$. Then $\{h, f\} = 0$ as well, i.e., the function $H$ is constant along the trajectories of the $g$-gradient system $\nabla f$. Therefore, the vector field $\nabla f$ is tangent to the level surface $H = E$. Since vector fields tangent to a given surface form a subalgebra of the Lie algebra of all vector fields (see 8.3.2), one can speak about the Lie algebra of integrals of the Hamiltonian system on a given energy level $H = E$.

EXAMPLE. In Kepler's problem the Hamiltonian $H(x,p)$ in $\mathbb{R}^6$ has the form

$$H(x,p) = \frac{p^2}{2m} + \frac{\alpha}{|x|}, \quad \alpha < 0.$$

In view of spherical symmetry there are three moment integrals:

$$M = (M_1, M_2, M_3) = [x,p] = \text{const}.$$

Besides, in this problem there are three more integrals:

$$W = (W_1, W_2, W_3) = \left[\frac{p}{m}, M\right] + \frac{\alpha x}{|x|} = \text{const},$$

which form the co-called *Laplace–Runge–Lenz vector.* Let us find the Poisson brackets of these integrals. First we check that

$$\{p_i, M_j\} = \varepsilon_{ijk} M_k, \quad \{M_i, M_j\} = \varepsilon_{ijk} M_k.$$

Taking into account these relations and the Leibniz rule, we obtain by direct calculations that

$$\{M_i, W_j\} = \varepsilon_{ijk} W_k.$$

It can be shown by some more tedious calculations that

$$\{W_i, W_j\} = -\frac{2E}{m} M_k.$$

Therefore, on the space of the first integrals generated by the functions $M_i, W_j$, $i,j = 1,2,3$, we have the Poisson algebra structure that depends on the energy $E$.

Consider this situation more closely from the point of view of Poisson algebras.

Even if the initial Poisson structure on the manifold $M$ was nondegenerate, the centralizer $Z_H$ of an element $H$, being a Poisson algebra, certainly has a nontrivial Casimir function, since

$$\{H, Z_H\} = 0$$

by definition. Now we define the *reduction of Poisson algebras.*

In the algebra $Z_H$ the function $H \in Z_H$ behaves as a constant. If we simply set $H = E$, this will give rise to the quotient Poisson algebra $Z_E$, where $E$ is a number:

$$R_E \colon Z_H \to Z_E, \quad \mathbb{Z}_H \subset C^\infty(M, \mathbb{R}).$$

We will call the homomorphism $R_E$ the *reduction homomorphism* of the initial system with Hamiltonian $H$ on an arbitrary Poisson manifold $M$. Now we make the following assumptions:

a) The Poisson structure on the $n$-dimensional manifold $M$ is nondegenerate, $\det(g^{ij}) \neq 0$, i.e, $(M, g)$ is a symplectic manifold, $n = 2l$.

b) The system is *topologically integrable* about a point $y \in M$, or, more precisely, about the trajectory of the system with Hamiltonian $H$ passing through this point. Formally, this means that there exist $C^\infty$-functions $z_1, \ldots, z_{n-2}$ such that

1) $\{z_j, H\} = 0$;

2) the equations $z_j = 0$, $j = 1, \ldots, n-2$, $H = E$ determine a nondegenerate trajectory of the initial system with Hamiltonian $H$.

The system is said to be *factorizable* globally on the level $H = E$ if condition b) if fulfilled everywhere on this level, i.e., the system is topologically integrable about any point $y \in M_E$.

We define the *ring of functions of the quotient space* (i.e., of the reduced phase space) to be the Poisson algebra $Z_E = C^\infty(M_{\text{Red}}; \mathbb{R})$ of smooth functions on the quotient space $M_{\text{Red}}$.

Local coordinates about a point of the quotient space

$$\{H = E\} = M_E \to M_{\text{Red}}, \quad y \to \bar{y},$$

are constructed as follows: these are functions $z_1, \ldots, z_{n-2}$ in condition b), which specify the trajectory passing through the point $y \in M_E \subset M$. Their images $(\bar{z}_j)$ in the quotient ring $Z_E$ are taken for the local coordinates of the point $\bar{y} \in M_{\text{Red}}$.

**13.1.6. Examples of Poisson algebras.** Here we give the fundamental examples of Poisson algebras.

1. The Poisson structure generated by the standard 2-form

$$\Omega = \sum dp_\alpha \wedge dx^\alpha$$

on the co-tangent manifold $T^*M$ with the Poisson bracket indicated above. For local canonical coordinates $y = (x^\alpha, p_\alpha)$ we have

$$\{x^\alpha, x^\beta\} = \{p_\alpha, p_\beta\} = 0, \quad \{x^\alpha, p_\beta\} = \delta^\alpha_\beta.$$

For a given Hamiltonian $H(x,p) = H(y)$ the equations of motion take the classical Hamiltonian form

$$\dot{x}^\alpha = \{x^\alpha, H\} = \frac{\partial H}{\partial p_\alpha},$$
$$\dot{p}_\alpha = \{p_\alpha, H\} = -\frac{\partial H}{\partial x^\alpha}$$

in view of the properties of the Poisson brackets. For any smooth function $f$, its directional derivative in the direction of the Hamiltonian field satisfies the *Liouville equation*

$$\dot{f} = \{f, H\} = \frac{\partial f}{\partial x^\alpha}\frac{\partial H}{\partial p_\alpha} - \frac{\partial f}{\partial p_\alpha}\frac{\partial H}{\partial x^\alpha}.$$

**Lemma 13.12.** *To any vector field $\xi(x)$ on a configuration space $N$ corresponds a unique Hamiltonian system on $T^*N$ with Hamiltonian $H = \xi^\alpha(x)p_\alpha$ and the standard symplectic structure.*

**Proof.** This follows trivially from the Hamilton equations

$$\dot{x}^\alpha = \frac{\partial H}{\partial p_\alpha} = \xi^\alpha(x),$$
$$\dot{p}_\alpha = -\frac{\partial H}{\partial x^\alpha} = p_\gamma \frac{\partial \xi^\gamma}{\partial x^\alpha}.$$

□

2. Now we consider the nonstandard symplectic (Poisson) structure on $T^*N = M$ specified by a magnetic field $B = (B_{\alpha\beta}(x))$ by means of the 2-form

$$\Omega = \sum_\alpha dp_\alpha \wedge dx^\alpha - \frac{e}{c}B,$$

where $e$ is the charge of the particle and $c$ the velocity of light in vacuum.

For the corresponding Poisson brackets we obtain

$$\{x^\alpha, x^\beta\} = 0, \quad \{x^\alpha, p_\beta\} = \delta^\alpha_\beta,$$
$$\{p_\alpha, p_\beta\} = \frac{e}{c}B_{\alpha\beta}(x).$$

The equations of motion of a particle with Hamiltonian $H(x,p)$ have the form

$$\dot{x}^\alpha = \{x^\alpha, H\}, \quad \dot{p}_\alpha = \{p_\alpha, H\}$$

and coincide with the equations in which a magnetic field independent of time is additionally included. This has been already discussed above.

Before we proceed to the next example, we will prove the *generalized Darboux theorem.* Its particular case for nondegenerate brackets has been already stated without proof.

**Theorem 13.8.** *For any Poisson structure $g^{ij}$ on a manifold $M$ having a constant rank in a neighborhood of a point $y_0 \in M$, there exist local coordinates $x^1, p_1, \ldots, x^n, p_n, z^1, \ldots, z^m$ in a neighborhood of this point such that*

$$\{x^i, x^j\} = \{p_i, p_j\} = 0, \quad \{p_i, x^j\} = \delta^i_j, \quad \{z_q, f\} = 0$$

*for all smooth functions $f$ on $M$.*

**Proof.** As the first step, we construct a pair of functions $x^1, p_1$ defined in a neighborhood of $y_0$ and such that $\{p_1, x^1\} = 1$. The Poisson tensor $g^{ij}$ does not vanish near the point $y_0$; hence there exists a function $x^1$ generating a Hamiltonian vector field that does not vanish at $y_0$. We take the function $p_1$ to be the time coordinate of this Hamiltonian system with reversed sign. By definition, we have $\{p_1, x^1\} = -1$, since $(-p_1)$ is the time and $x^1$ the Hamiltonian:

$$\dot{p}_1 = \{p^1, x_1\} = -1.$$

At the second step we consider the local action of the group $\mathbb{R}^2$ generated by the Hamiltonians $x^1$ and $p_1$. The vector fields $\nabla x^1$ and $\nabla p_1$ commute with each other since the commutator of these fields is generated by the Poisson bracket of the functions $\{p_1, x^1\}$, which is equal to one, so that the commutator is equal to zero. We choose the local coordinates $x^1, p_1, w^1, \ldots, w^{m-2}$, $m = \dim M$, by the existence theorem for a pair of commuting ordinary differential equations. Here the $w^j$ by definition are integrals of action of $\mathbb{R}^2$. Therefore, we have

$$\{w^j, x^1\} = \{w^j, p_1\} = 0, \quad j = 1, \ldots, m-2.$$

At the third step we apply this procedure to coordinates $w$ and repeat it until the remaining Poisson brackets $\{w^i, w^j\}$ become equal to zero at $y_0$. This completes the proof for brackets of locally constant rank. □

This theorem immediately implies the following result.

**Corollary 13.2.** *For Poisson brackets of locally constant rank, the set of Casimir functions is locally well defined: their levels generate a symplectic foliation for which the restrictions of the Poisson brackets to leaves are nondegenerate.*

This corollary follows directly from the theorem by taking the set of local Casimir functions to be the local coordinates $z^j$.

3. Now we consider the case where the configuration space $N$ is a Lie group $G = N$. For example, as we pointed out before, $G = \mathrm{SO}(3)$ is the configuration space for the motion of a solid body about a fixed point. The tangent manifold $(g, \dot{g}) \in TN$ is represented as the Cartesian product

$$TN = G \times \mathbb{R}^n$$

for $N = G$, $\dim G = n$. This isomorphism is established by left translation from the point $1 \in G$ to any point $g_0 \in G$:

$$L_{g_0}: G \to G, \quad L_{g_0}(g) = g_0 g,$$

where the differential $dL_{g_0}$ establishes an isomorphism between the tangent space at the point $g = g_0$ and the Lie algebra realized as the tangent space at the point 1. In particular, for a curve $g(t) \in G$, the vector $g^{-1}(t)\dot{g}(t)$ lies in the Lie algebra (realized as the tangent space at $g_0 = 1$). Any Lagrangian $\tilde{L}(g, \dot{g})$ on the group $G$ can be conveniently written in variables $(g, g^{-1}\dot{g}) = (g, A)$, where $A$ is an element of the Lie algebra $\mathfrak{g}$.

Let the Lagrangian be written in these variables as $L(g, g^{-1}\dot{g}) = L(g, A)$. Then it is said to be *left-invariant* if $\tilde{L} = \tilde{L}(A)$.

Let $e_1, \dots, e_n$ be a basis of the Lie algebra $\mathfrak{g}$; thus any of its elements can be written as $x^\alpha e_\alpha$. In a neighborhood of the identity $1 \in G$, the exponential mapping determines the coordinates

$$g = \exp(x^\alpha e_\alpha) \leftrightarrow g = (x^1, \dots, x^n),$$

where $\exp(A) = \sum_{n \geq 0} A^n/n!$ for matrix groups, and $A$ is the tangent matrix (these are coordinates of the first kind; see 6.1.2).

The expression $g^{-1}\dot{g}$ is the generic symbol for all left-invariant fields on the group $G$: indeed,

$$L_{g_0}(g^{-1}\dot{g}) = (g_0 g)^{-1}(g_0 g)^{\cdot},$$

where $g_0 = \text{const}$. Hence we have

$$(g_0 g)^{-1}(g_0 g)^{\cdot} = g^{-1} g_0^{-1} g_0 \dot{g} = g^{-1}\dot{g}.$$

We assume that the elements $e_\alpha \in \mathfrak{g}$ are extended to left-invariant fields on the group $G$ (to be denoted by the same symbols) and the $M_\alpha$ are Hamiltonians in $T^*G = G \times \mathbb{R}^n$ corresponding to these fields (see Lemma 13.12). By the definition of the $M_\alpha$, their pairwise Poisson brackets are linear functions of them:

$$\{M_\alpha, M_\beta\} = C_{\alpha\beta}^\gamma M_\gamma.$$

For the variables $x^\alpha$ we obviously have

$$\{x^\alpha, x^\beta\} = 0.$$

If we regard

$$e_\alpha = \psi_\alpha^\gamma(x)\,\frac{\partial}{\partial x^\gamma}$$

as left-invariant fields on the group $G$, then in canonical coordinates $(x,p)$ on $T^*G$ about the point $g=1$, where $g=\exp(x^\alpha e_\alpha)$, we have by definition

$$\{x^\alpha, x^\beta\} = \{p_\alpha, p_\beta\} = 0, \quad \{x^\alpha, p_\beta\} = \delta_\beta^\alpha,$$
$$M_\alpha = \psi_\alpha^\gamma(x)p_\gamma.$$

The Poisson brackets have the form

$$\{M_\alpha, x^\beta\} = \{\psi_\alpha^\gamma(x)p_\gamma, x^\beta\} = -\psi_\alpha^\gamma(x)\delta_\gamma^\beta = -\psi_\alpha^\beta(x).$$

By the definition of the fields $e_\alpha$, the normalizing condition

$$\psi_\alpha^\beta(0) = \delta_\alpha^\beta$$

holds at $g=1$.

Taking into account the type of coordinates $x^\beta$, where $g=\exp(x^\alpha e_\alpha)$, we see that the matrix functions $\psi_\alpha^\beta(x)$ are determined by the group $G$ and the Poisson tensor is

$$g^{ij} = \begin{pmatrix} 0 & \psi(x) \\ -\psi^t(x) & C_{\alpha\beta}^\gamma M_\gamma \end{pmatrix} = g^{ij}(x,M).$$

For left-invariant Lagrangian (Hamiltonian) systems the Hamiltonian has the form

$$H = H(M_1,\ldots,M_n).$$

Hence we obtain a closed equation

$$\dot M_\alpha = \frac{\partial H}{\partial M_\beta}\{M_\beta, M_\alpha\} = \frac{\partial H(M)}{\partial M_\beta}\cdot C_{\beta\alpha}^\gamma M_\gamma.$$

The quantity $\partial H/\partial M_\beta$ is called the *angular velocity* $\Omega^\beta = \partial H/\partial M_\beta$:

$$\dot M_\alpha = \Omega^\beta M_\gamma C_{\beta\alpha}^\gamma.$$

For compact Lie groups there is a unique (up to proportionality) Killing metric. This is a Riemannian (i.e., positive definite) metric that is invariant under left and right translations. Using index raising and lowering in this metric, we can write the Hamilton equation as

$$\dot M^\alpha = \Omega^\beta M^\gamma C_{\gamma\beta\alpha}.$$

For compact Lie groups one can find a basis in the Lie algebra $\mathfrak{g}$ such that the Killing metric will be $\delta_{\alpha\beta}$. In this case one can show that the tensor $C_{\gamma\beta\alpha}$ is skew-symmetric relative to permutations of subscripts $(\alpha,\beta,\gamma)$ (e.g., for SO(3) this is the tensor $C_{ijk} = \varepsilon_{ijk}$, and the proof in the general case reduces to this one by imbedding the compact group into SO$(n)$). In the

particular case under consideration, the right-hand side of the Hamilton equation becomes the commutator in the Lie algebra:

$$\dot{M} = [\Omega, M], \quad \Omega = \frac{\partial H}{\partial M}.$$

This is a *generalization of Euler's equations* for a free solid body.

For the group $G = \mathrm{SO}(n)$ the Hamiltonian is

$$H = \sum_j \frac{M_{ij}^2}{q_i + q_j},$$

where $i, j = 1, 2, \ldots, n$, $i < j$, $q_j > 0$. The group $G = \mathrm{E}(3)$ of motions of the space $\mathbb{R}^3$ arises in connection with Kirchhoff equations for the motion of a solid body in a fluid.

Our discussion leads us to an important example of Poisson algebras.

Consider the manifold $M = \mathfrak{g}^* = \mathbb{R}^n$, which is the linear space dual to the Lie algebra $\mathfrak{g} = \mathbb{R}^n$, and let $M_1, \ldots, M_n$ be a basis in the space of linear functions on $M$ (i.e., in the Lie algebra $\mathfrak{g}$) with Poisson bracket

$$\{M_\alpha, M_\beta\} = C_{\alpha\beta}^\gamma M_\gamma$$

(here $C_{\alpha\beta}^\gamma$ are structural constants of the algebra $\mathfrak{g}$). This bracket can be extended to the Poisson bracket on the space of all smooth functions on $M = \mathfrak{g}^*$. It is called the *Lie–Poisson bracket*, and its restriction to symplectic leaves is called the *Kirillov form.*

This bracket has Casimir functions. Consider the *universal enveloping algebra* $U(\mathfrak{g})$ of the Lie algebra $\mathfrak{g}$. This is the algebra with addition and associative multiplication generated by the elements $1, M_1, \ldots, M_n$ with relations

$$M_\alpha M_\beta - M_\beta M_\alpha = C_{\alpha\beta}^\gamma M_\gamma.$$

*Casimir functions of this algebra* are defined to be polynomials $K$ in generators $M_1, \ldots, M_n$ such that

$$[K, M_\alpha] = 0, \quad \alpha = 1, \ldots, n.$$

Although the algebra is not commutative, any element $K \in U(\mathfrak{g})$ can be written as a linear combination of monomials of the form $M_{i_1}^{m_1} \cdots M_{i_k}^{m_k}$, where $i_1 < i_2 < \cdots < i_k$. The quantity $m = m_1 + \cdots + m_k$ is called the degree of such an ordered monomial. For the element $K$ we have

$$K = K_m + \cdots + K_0,$$

where the degree of $K_j$ is $j$.

**Lemma 13.13.** *If* $[K, M_\alpha] = 0$ *for all* $\alpha = 1, \ldots, n$ *in the algebra* $U(\mathfrak{g})$, *then the following equalities hold in the Poisson algebra:*

$$\{K_m, M_\alpha\} = 0, \quad \alpha = 1, \ldots, n.$$

**Proof.** The Poisson bracket of any homogeneous polynomial of degree $q$ with elements $M_\alpha$ is again a polynomial of degree $q$,

$$\deg\{K_q, M_\alpha\} = q,$$

by the definition of a linear bracket

$$\{M_\alpha, M_\beta\} = C^\gamma_{\alpha\beta} M_\gamma$$

and the Leibniz identity.

The commutator of any ordered homogeneous polynomial with $M_\alpha$ is

$$[K_q, M_\alpha] = \{K_q, M_\alpha\} + K'_{q-1},$$

where the remainder $K_{q-1}$ is of lower degree. Therefore, the condition $[K, M_\alpha] = 0$ implies that $\{K_m, M_\alpha\} = 0$. Hence the lemma. □

CONCLUSION. The set of Casimir functions of the Lie–Poisson brackets for any Lie algebra $\mathfrak{g}$ is a ring generated by homogeneous polynomials. If in the enveloping algebra $U(\mathfrak{g})$ we choose a basis consisting of Casimir functions relative to the commutator that are homogeneous polynomials, then there is a one-to-one correspondence between the Casimir functions of the enveloping algebra $U(\mathfrak{g})$ and those of the Poisson algebra on $\mathfrak{g}$. We will not prove here the possibility of the choice of a homogeneous basis of Casimir functions in $U(\mathfrak{g})$.

**13.1.7. Canonical transformations.** Consider an arbitrary symplectic manifold. It will be convenient to write Hamiltonian systems in canonical coordinates.

**Theorem 13.9.** *If the Hamilton function $H(x,p)$ does not depend explicitly on time, then the form $\Omega = \sum_{i=1}^n dx^i \wedge dp_i$ is preserved along the Hamiltonian system:*

$$\dot{\Omega} = 0.$$

**Proof.** We will use the following facts (see 8.3.6):

$$\frac{d}{dt}(\Omega_1 \wedge \Omega_2) = \frac{d}{dt}\Omega_1 \wedge \Omega_2 + \Omega_1 \wedge \frac{d}{dt}\Omega_2, \tag{13.7}$$

$$\frac{d}{dt}(dp_i) = -d\Big(\frac{\partial H}{\partial x^i}\Big) = -\frac{\partial^2 H}{\partial x^i \partial x^j}\, dx^j - \frac{\partial^2 H}{\partial x^i \partial p_j}\, dp_j,$$

$$\frac{d}{dt}(dx^i) = d\Big(\frac{\partial H}{\partial p_i}\Big) = \frac{\partial^2 H}{\partial p_i \partial x^j}\, dx^j + \frac{\partial^2 H}{\partial p_i \partial p_j}\, dp_j.$$

Therefore

$$\frac{d}{dt}\Big(\sum_i dp_i \wedge dx^i\Big) = \sum_{i,j}\Big\{-\frac{\partial^2 H}{\partial x^i \partial x^j}\, dx^j \wedge dx^i - \frac{\partial^2 H}{\partial x^i \partial p_j}\, dp_j \wedge dx^i + dp_i \wedge \frac{\partial^2 H}{\partial p_i \partial x^j}\, dx^j + dp_i \wedge \frac{\partial^2 H}{\partial p_i \partial p_j}\, dp_j\Big\} = 0,$$

since the exterior product is skew-symmetric. Hence the theorem. □

**Corollary 13.3** (Liouville's theorem). *A Hamiltonian system preserves the phase volume of any domain in the $(x, p)$-space.*

**Proof.** By Lemma 13.2, the volume element of a $2n$-dimensional symplectic manifold is $\Omega^n = \Omega \wedge \cdots \wedge \Omega$. Since $\dot{\Omega} = 0$, formula (13.7) implies that $(\Omega \wedge \cdots \wedge \Omega)^{\cdot} = 0$. Hence the corollary. □

A transformation $\Phi\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ of the phase space $(x, p)$ (or, more generally, of a symplectic manifold) is said to be *canonical* if it preserves the skew-symmetric scalar product (i.e., if the form $\Omega$ transforms into itself).

In other words, canonical transformations are "motions" of the skew-symmetric "metric". By Theorem 13.9, translations along the trajectories of the Hamiltonian system with Hamiltonian $H(x, p)$ give rise to a one-parameter group of canonical transformations. We may assume here that the Hamiltonian is defined only locally, i.e., only the closed 1-form $dH$ is well defined globally on the manifold $M$. The converse is also true.

**Theorem 13.10.** *Let $\Phi_t(x, p) = (x(t), p(t))$ be a local one-parameter group of canonical transformations and $X = \frac{d\Phi_t}{dt}\big|_{t=0}$ the corresponding vector field.*

*Then there is a locally single-valued smooth Hamilton function $H(x, p)$ (i.e., a closed 1-form $dH$) relative to which $X$ is a Hamiltonian vector field, i.e.,*

$$X = \Big(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial x}\Big).$$

**Proof.** Let $X = (A^i, B_i)$, $i = 1, \ldots, n$, in coordinates $(x, p)$. Consider the translation by a small time $\Delta t$. We have

$$x^i \to x^i + A^i(x, p)\Delta t + O(\Delta t^2) = \tilde{x}^i,$$
$$p_i \to p_i + B_i(x, p)\Delta t + O(\Delta t^2) = \tilde{p}_i.$$

The skew-symmetric metric is preserved:

$$\sum_i d\tilde{p}_i \wedge d\tilde{x}^i = \sum_i dp_i \wedge dx^i.$$

Therefore, we have

$$\begin{aligned} d\tilde{p}_i \wedge d\tilde{x}^i &= (dp_i + (dB_i)\Delta t) \wedge (dx^i + (dA^i)\Delta t) \\ &= dp_i \wedge dx^i + \Delta t \Big[ \frac{\partial B_i}{\partial p_j}\, dp_j \wedge dx^i + \frac{\partial B_i}{\partial x^j}\, dx^j \wedge dx^i \\ &\qquad + \frac{\partial A^i}{\partial x^j}\, dp_i \wedge dx^j + \frac{\partial A^i}{\partial p_j}\, dp_i \wedge dp_j \Big] + O(\Delta t^2). \end{aligned}$$

Hence

$$\begin{aligned} \frac{\partial A^i}{\partial p_j}\, dp_i \wedge dp_j = 0 &\iff \frac{\partial A^i}{\partial p_j} = \frac{\partial A^j}{\partial p_i}, \\ \frac{\partial B_i}{\partial x^j}\, dx^i \wedge dx^j = 0 &\iff \frac{\partial B_i}{\partial x^j} = \frac{\partial B_j}{\partial x^i}, \\ \frac{\partial A^i}{\partial x^j}\, dp_i \wedge dx^j + \frac{\partial B_i}{\partial p_j}\, dp_j \wedge dx^i = 0 &\iff \frac{\partial A^i}{\partial x^j} = -\frac{\partial B_j}{\partial p_i}. \end{aligned}$$

These equalities are equivalent to the property that the form

$$\omega = A^i\, dp_i - B_i\, dx^i$$

is closed, $d\omega = 0$. But locally each closed form is exact:

$$\omega = dH, \quad A^i = \frac{\partial H}{\partial p_i}, \quad B_i = -\frac{\partial H}{\partial x^i}.$$

Hence the theorem. □

If a Hamiltonian $H(x,p,t)$ explicitly depends on time, then it specifies a one-parameter family of canonical transformations $\varphi_t$, which are translations by time $t$ along the trajectories, but this family is not a group: $\varphi_{t_1+t_2} \neq \varphi_{t_1} \circ \varphi_{t_2}$.

For $n = 1$ the canonical transformations are precisely the ones that preserve the area, i.e., the form $\Omega = dx \wedge dp$.

For $n > 1$ there are more volume-preserving than canonical transformations.

This can be seen already for linear transformations. In canonical coordinates $(x^1, \dots, x^n, p_1, \dots, p_n)$, the volume-preserving transformations constitute the group $\mathrm{SL}(2n)$, whereas the linear canonical transformations are precisely the symplectic transformations, which form the group $\mathrm{Sp}(n, \mathbb{R})$.

Note that the Lie algebra of the group $\mathrm{Sp}(n, \mathbb{R})$ is isomorphic to the Lie algebra of quadratic potentials relative to the Poisson bracket. Indeed, let the Hamiltonian $H(x,p)$ be a quadratic form:

$$H = \frac{1}{2} \sum_{i,j=1}^{n} (a_{ij} x^i x^j + 2 b_{ij} x^i p_j + c_{ij} p_i p_j) = \frac{1}{2} (x \;\; p) \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix},$$

where $a_{ij} = a_{ji}$, $c_{ij} = c_{ji}$. For example, so is the potential of a *harmonic oscillator*

$$H(x,p) = \frac{1}{2m} p^2 + m\omega^2 x^2,$$

where $\omega$ is the frequency. For a quadratic Hamiltonian, the Hamilton equations become

$$\dot{y} = Ky, \quad K = \begin{pmatrix} B^\top & C \\ -A & -B \end{pmatrix}, \tag{13.8}$$

where $y = (x,p)$. The Lie algebra of the group $\mathrm{Sp}(n, \mathbb{R})$ consists of matrices (13.8), and the Lie algebra of such matrices is isomorphic to the Lie algebra of quadratic potentials relative to the Poisson bracket.

## 13.2. Lagrangian submanifolds and their applications

**13.2.1. The Hamilton–Jacobi equation and bundles of trajectories.** Let $M$ be a phase space with coordinates $p_i, x^i$, $i = 1, \dots, n$, i.e., a symplectic manifold with a closed form $\Omega = \sum_i dp_i \wedge dx^i$.

A submanifold $\Gamma \subset M$ of dimension $n$ is said to be *Lagrangian* or an $n$-dimensional *Lagrangian surface* if the restriction to it of the symplectic form $\Omega$ is equal to zero:

$$\Omega|_\Gamma \equiv 0.$$

This is equivalent to the property that at each point $x \in \Gamma$ the tangent $n$-plane to $\Gamma$ is Lagrangian in the tangent space to $M$ (see 2.1.2).

EXAMPLE. Any Lagrangian subspace in $\mathbb{R}^{2n}$ with symplectic product specified by the form $\sum_i dp_i \wedge dx^i$ (see 2.1.2) is Lagrangian. This is true, e.g., for the $n$-planes

$$x^1 = \dots = x^n = 0 \quad \text{or} \quad p_1 = \dots = p_n = 0.$$

Now we will give other (equivalent) definitions of Lagrangian surfaces (submanifolds) in the phase space and extended phase space.

A submanifold $\Gamma^n$ of the phase space with canonical coordinates $(x,p)$ is said to be *Lagrangian* if in a small neighborhood $U$ of any point of $\Gamma^n$, the truncated action (12.19)

$$S_0 = \int_\gamma p\, dx$$

does not depend on the path $\gamma \subset U$ in $\Gamma^n$, being a function of its endpoints.

A submanifold $\Gamma^{n+1}$ of the extended phase space with canonical coordinates $(x, p, E, t)$ is said to be *Lagrangian* if locally in a neighborhood of any

point of $\Gamma^{n+1}$, the action (12.18)

$$S_0 = \int_\gamma (p\,dx - E\,dt)$$

is a function of the endpoints of the path.

The equivalence of these definitions obviously follows from the Stokes formula. Indeed, the actions are functions of the endpoints if the integrands are closed forms on $\Gamma$, which means that

$$\Big(\sum_i dp_i \wedge dx^i\Big)\Big|_{\Gamma^n} \equiv 0 \quad \text{or} \quad \Big(\sum_i dp_i \wedge dx^i - dE \wedge dt\Big)\Big|_{\Gamma^{n+1}} \equiv 0.$$

**Lemma 13.14.** 1. *A submanifold $\Gamma^n$ specified in the phase space as the graph of a mapping $p_i = f_i(x)$, $i = 1, \dots, n$, is Lagrangian if and only if it can be written locally (in a neighborhood of each of its points) as*

$$p_i = \frac{\partial S_0}{\partial x^i}, \quad S_0 = S_0(x^1, \dots, x^n). \tag{13.9}$$

2. *A submanifold $\Gamma^{n+1}$ specified in the extended phase space as the graph of a mapping $p_i = f_i(x,t)$, $i = 1, \dots, n$, $E = E(x,t)$, is Lagrangian if and only if it is locally given by the formulas*

$$p_i = \frac{\partial S}{\partial x^i}, \quad E = -\frac{\partial S}{\partial t}, \quad S = S(x^1, \dots, x^n, t).$$

**Proof.** If the submanifold $\Gamma^n$ is Lagrangian, then we can take for $S_0$ the truncated action with $x^1, \dots, x^n$ treated as coordinates on $\Gamma^n$. Conversely, if $\Gamma^n$ is as in (13.9), then

$$\Big(\sum_i p_i\,dx^i\Big)\Big|_{\Gamma^n} = \sum_i \frac{\partial S_0}{\partial x^i}\,dx^i = dS_0$$

and $d^2 S_0 = \big(\sum_i dp_i \wedge dx^i\big)\big|_{\Gamma^n} \equiv 0$.

A similar proof works for submanifolds $\Gamma^{n+1}$ in the extended phase space. □

The relation $E = H(x,p,t)$ implies that in the extended phase space, the action function satisfies the equation

$$\frac{\partial S}{\partial t} + H\Big(x, \frac{\partial S}{\partial x}, t\Big) = 0,$$

which is called the *Hamilton–Jacobi equation.*

**Lemma 13.15.** *Under canonical transformations, Lagrangian submanifolds map into Lagrangian submanifolds.*

**Proof.** If $F\colon M \to M$ is a canonical transformation and $\Gamma \subset M$ is a Lagrangian submanifold, then

$$\Omega(F_*\xi, F_*\eta) = F^*\Omega(\xi, \eta) = \Omega(\xi, \eta) = 0.$$

Hence the lemma. □

Now we turn to the most important example of Lagrangian surfaces.

Let $\Gamma$ be a Lagrangian surface in the phase space $\mathbb{R}^{2n}$. Consider a Hamiltonian system with Hamiltonian $H(x,p,t)$ and denote by $\hat{\Gamma}$ the *bundle of trajectories*, i.e., the submanifold in the extended phase space $\mathbb{R}^{2n+2}$ consisting of all trajectories intersecting $\Gamma$. We may assume that they intersect $\Gamma$ for $t = 0$. Any section of the surface $\hat{\Gamma}$ at time $t =$ const is the translation $\Gamma_t$ of the initial surface $\Gamma$ by time $t$ along the trajectories of the flow.

**Lemma 13.16.** *The submanifold $\hat{\Gamma}$ is Lagrangian.*

**Proof.** It suffices to show that the restriction of the form $\sum_i p_i\, dx^i - E\, dt$ to this submanifold is closed. Since $\hat{\Gamma}$ is a bundle of trajectories, $E = H(x,p,t)$ on $\hat{\Gamma}$.

Consider the symplectic form $\hat{\Omega} = \sum_i dp_i \wedge dx^i - dE \wedge dt$ in the extended phase space. For $t = t_0$ the tangent space at the point $(x, p, E, t_0) \in \hat{\Gamma}$ is generated by the vectors

$$v = \left(\xi^1, \dots, \xi^n, \eta_1, \dots, \eta_n, \frac{\partial H}{\partial x^i}\xi^i + \frac{\partial H}{\partial p_i}\eta_i, 0\right)$$

(the tuples of vectors $(\xi, \eta)$ are not arbitrary) and the vector

$$w = \left(\dot{x}^1, \dots, \dot{x}^n, \dot{p}_1, \dots, \dot{p}_n, \frac{\partial H}{\partial t}, 1\right)$$

tangent to the trajectory. Obviously, for any two such vectors $v, v'$ we have $\hat{\Omega}(v, v') = 0$, since the submanifold $\Gamma$ is Lagrangian in the phase space. By virtue of the Hamilton equations, $\hat{\Omega}(v, w) = 0$ for any tangent vector $v$ (which is verified by a direct substitution). Therefore, $\hat{\Omega}|_{\hat{\Gamma}} = 0$ at the points of $\Gamma$.

A symplectic form $\hat{\Omega}$ is preserved by the Hamiltonian flow. The proof of this fact is similar to that of Theorem 13.9, differing in that one has to consider the extended phase space with form $\hat{\Omega}$ and Hamiltonian $\hat{H}(x,p,t,E) = H(x,p,t) - E$.

Since the form $\hat{\Omega}$ is conserved by the flow and its restriction to $\hat{\Gamma}$ equals zero for $t = 0$, it vanishes everywhere. Hence the lemma. □

As an example of such a Lagrangian submanifold, consider the bundle of trajectories originating at the surface $\Gamma = (x_0, p)$ (the point $x_0$ of the configuration space is fixed, while the momenta $p$ are arbitrary). For $t = 0$

the surface $\Gamma$ cannot be specified as the graph of a mapping $p = \frac{\partial S}{\partial x}$, but for any small $t > 0$, its translation by time $t$ can. We will consider such bundles in Section 13.2.3.

Now we indicate an important consequence of the above results. Suppose a Hamilton–Jacobi equation

$$\frac{\partial S}{\partial t} + H\Big(x, \frac{\partial S}{\partial x}, t\Big) = 0$$

is given and we look for a solution to this equation with initial condition

$$S(x^1, \dots, x^n, 0) = f(x^1, \dots, x^n).$$

We construct the bundle of trajectories starting at $t = 0$ from the Lagrangian surface in $\mathbb{R}^{2n}$ specified as the graph of the function

$$p_i = \frac{\partial f}{\partial x^i}(x^1, \dots, x^n).$$

Suppose for small times $t$ the surfaces $\Gamma_t$ are diffeomorphically projected to the $x$-plane in $\mathbb{R}^n$. Then with each point $x$ of $\mathbb{R}^n$ we can associate its inverse image $P(x,t)$ and set

$$S(x,t) = f(x) + \int_{P(x,0)}^{P(x,t)} (p_i\, dx^i - H\, dt).$$

The function $S(x,t)$ provides the required solution to the Hamilton–Jacobi equation.

Suppose the Hamiltonian $H = H(x,p)$ does not explicitly depend on time. Then the following theorem holds.

**Theorem 13.11.** 1. *The vector* $\nabla H = \left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial x}\right) = (\dot{x}, \dot{p})$ *is tangent to the surface* $H = E_0 = \text{const}$ *and its (skew-symmetric) scalar product with any tangent vector to this surface equals zero.*

2. *At any point of an $n$-dimensional Lagrangian surface $\Gamma$ lying on the energy level $H(x,p) = E_0$, the vector $\nabla H$ is a tangent vector; in particular, if a trajectory of the Hamiltonian system with Hamiltonian $H$ is tangent to the surface $\Gamma$, then it entirely belongs to this surface.*

**Proof.** 1. A vector $\xi$ is tangent to the surface $H = E_0$ if and only if $\xi^i \frac{\partial H}{\partial y^i} = 0$, $y = (x,p)$. We have

$$\langle \xi, \nabla H \rangle = \xi^i g_{ij} (\nabla H)^j = \xi^i g_{ij} g^{jk} \frac{\partial H}{\partial y^k} = \xi^i \frac{\partial H}{\partial y^i} = 0.$$

2. A symplectic product is nondegenerate, and if the symplectic scalar product of some vector with any vector of a Lagrangian plane is zero, then this vector lies in this plane. Since $\langle \nabla H, \xi_i \rangle = 0$ for all basis vectors $\xi_1, \dots, \xi_n$ in the tangent plane at any point of $\Gamma$, the vector $\nabla H$ itself is tangent to $\Gamma$.

The integral curves of the field $\nabla H$ tangent to $\Gamma$ at least at one point lie entirely in $\Gamma$. The proof is completed. □

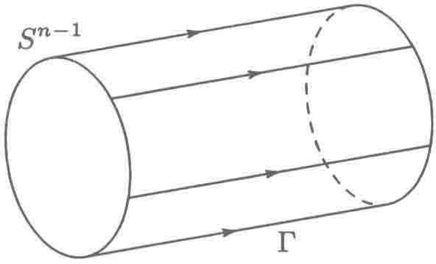


**Figure 13.1.** Bundle of trajectories on a level surface.

**Corollary 13.4.** *Let $S^{n-1}$ be an $(n-1)$-dimensional surface on the level $H(x,p) = E_0$ and let the restriction of the form $\Omega$ to it be equal to zero. Draw the trajectories of the Hamiltonian system in $(x,p)$-space through the points of this surface. Suppose that the resulting surface is n-dimensional. Then this n-dimensional surface $\Gamma^n$ is Lagrangian and lies on the energy level $E_0$. If the surface $\Gamma^n$ is specified (locally) by the equation*

$$p_i = f_i(x) = \frac{\partial S_0}{\partial x^i},$$

*then the function $S_0(x)$ satisfies the truncated Hamilton–Jacobi equation*

$$E_0 = H\Big(x, \frac{\partial S_0}{\partial x}\Big),$$

*and the function $S(x,t) = \int p\,dx - H\,dt$ is written as*

$$S(x,t) = S_0(x) - E_0 t, \quad \frac{\partial S}{\partial t} + H\Big(x, \frac{\partial S_0}{\partial x}\Big) = 0.$$

Note that the 1-form $\alpha = p_i\,dx^i - E\,dt$ restricted to the surface of energy level $H(x,p) = E$ in the extended phase space $\mathbb{R}^{2n+2}$ is not closed and satisfies everywhere the relation

$$\alpha \wedge (d\alpha)^n \neq 0.$$

Manifolds of dimension $(2n+1)$ endowed with a 1-form $\alpha$ satisfying such a relation are said to be *contact*, and the form $\alpha$ is called a *contact form* or a *contact structure*.

For contact structures, which locally are always obtained as restrictions of the form $p\,dx - E\,dt$ to an energy level surface, the following analog

of the Darboux theorem holds (and is proved using this very theorem): Any contact structure can be locally reduced in appropriate coordinates $(x^1, \dots, x^n, p_1, \dots, p_n, t)$ to the form

$$\alpha = \sum_i p_i \, dx^i - dt.$$

**13.2.2. Representation of canonical transformations.** Transitions to other canonical coordinates (i.e., canonical transformations) play a very important role in many analytic studies in the theory of Hamiltonian systems, as well as in applied computations, especially in celestial mechanics. There, one must consider not only one-parameter families specified by Hamiltonians $H$, but also finite canonical transformations. We will give a convenient form of their representation in canonical coordinates $(x, p)$ assuming that the transformation is sufficiently close to the identity.

Suppose we are given a mapping of symplectic manifolds,

$$\Phi \colon M \to M',$$

written locally as

$$y' = y'(y), \quad y' = (y'^1, \dots, y'^{2n}), \quad y = (y^1, \dots, y^{2m}).$$

Consider the Cartesian product

$$M \times M'$$

with local coordinates $(y, y')$ and 2-form $\Omega - \Omega' = \tilde{\Omega}$,

$$\tilde{\Omega} = \sum_{i<j} g_{ij}(y) \, dy^i \wedge dy^j - \sum_{k<l} g'_{kl}(y') \, dy'^k \wedge dy'^l.$$

**Theorem 13.12.** *The mapping $\Phi$ is symplectic, i.e., $\Phi^*\Omega' = \Omega$, if and only if the restriction of $\tilde{\Omega}$ to the graph of this mapping,*

$$M_\Phi = \{(y, \Phi(y))\} \subset M \times M',$$

*is equal to zero. For $m = n$ this means that $M_\Phi$ is a Lagrangian submanifold in $M \times M'$.*

**Proof.** We may assume that $m \le n$, since for $m > n$ the form $\Phi^*\Omega'$ is degenerate.

Let the mapping be symplectic. This simply means that the scalar product of vectors is preserved. If $\xi_1$, $\xi_2$ are tangent vectors to $M$ at the point $y$, and $\xi'_1 = d\Phi(\xi_1)$, $\xi'_2 = d\Phi(\xi_2)$ are the corresponding vectors at the point $\Phi(y) = y'$, then

$$\langle \xi'_1, \xi'_2 \rangle = \langle \xi_1, \xi_2 \rangle.$$

Any tangent vector to the graph $M_\Phi \subset M \times M'$ has the form $(\xi, \xi')$, where $\xi' = d\Phi(\xi)$ at the point $(y, \Phi(y)) \in M_\Phi$.

By the definition of the form $\tilde{\Omega} = \Omega - \Omega'$, for a pair of vectors, we have

$$\tilde{\Omega}((\xi_1, \xi_1'), (\xi_2, \xi_2')) = \Omega(\xi_1, \xi_2) - \Omega'(\xi_1', \xi_2'),$$

where $\xi_i' = d\Phi(\xi_i)$. Hence the theorem. □

Consider now the important case $m = n$ in canonical coordinates. Suppose we are given a canonical transformation $\Phi$:

$$x' = x'(x,p), \quad p' = p'(x,p),$$

$\Omega = dp \wedge dx$, $\Omega' = dp' \wedge dx'$, and $\Phi^*\Omega' = \Omega$. Let $M_\Phi$ be the graph of the transformation $\Phi$. By the above theorem, this is a Lagrangian submanifold in $M \times M' \supset M_\Phi$.

Suppose that $M = M'$ and the transformation $\Phi$ is close to the identity. In this case the graph $M_\Phi$ is close to the diagonal $(y, y)$. The Lagrangian submanifold $M_\Phi$ is projected (locally) to Lagrangian subspaces of the variables $(x, p')$ or $(p, x')$ with $\tilde{\Omega} = dp \wedge dx - dp' \wedge dx'$.

Case 1: $(x, p')$. Since the manifold $M_\Phi$ is Lagrangian, it can be locally specified as a graph: there is a function $S_1(x, p')$ such that $M_\Phi$ has the form

$$p_\alpha = \frac{\partial S_1}{\partial x^\alpha}, \quad x'^\alpha = \frac{\partial S_1}{\partial p'_\alpha}.$$

Case 2: $(p, x')$ (the "dual choice"). There is a function $S_2(p, x')$ such that $M_\Phi$ has the form

$$x^\alpha = \frac{\partial S_2}{\partial p_\alpha}, \quad p'_\alpha = \frac{\partial S_2}{\partial x'^\alpha}.$$

In view of the properties of the Legendre transformation

$$(x, p') \to (p, x'),$$

which has already been used in deriving the Hamilton equations from the Euler–Lagrange equations, we have

$$x^\alpha p_\alpha + x'^\alpha p'_\alpha - S_1(x, p') = S_2'(p, x').$$

Recall that for the form $d\eta \wedge d\xi$ we started with the function $S_1(\eta)$ and the submanifold

$$\xi = \frac{\partial S_1(\eta)}{\partial \eta}.$$

The same submanifold may be written (locally) as

$$\eta = \frac{\partial S_2(\xi)}{\partial \xi},$$

where $\eta\xi - S_1(\eta) = S_2(\xi)$ and the equation $\xi = S_1(\eta)$ is (locally) solvable. In our case $\Omega = dp \wedge dx - dp' \wedge dx'$, $\xi = (x, p')$, and $\eta = (p, x')$.

In some cases it is expedient to choose other Lagrangian projections, e.g., when the mapping $\Phi$ is not close to the identity.

**13.2.3. Conical Lagrangian surfaces.** Consider separately Hamiltonians $H(x,p)$ that are first-order homogeneous functions with respect to the momentum,

$$H(x,\lambda p)=\lambda H(x,p),\quad \lambda>0.$$

Such is, e.g., the Hamiltonian for rays of light in isotropic medium:

$$H(x,p)=c(x)|p|.$$

For such a Hamiltonian is suffices to know its trajectories on a single nonzero energy level $H = E_0$ (for example, $E_0 = 1$). All other trajectories are obtained from these by dilatation $p\to\lambda p$, $H\to\lambda H$.

EXAMPLE. GEODESICS. The Hamiltonian for geodesics of a metric $g_{ij}(x)$ is $H=g^{ij}p_ip_j$, but they can also be obtained from the Hamiltonian $H'=\sqrt{H}=\sqrt{g^{ij}p_ip_j}$. The corresponding Hamilton equations on each energy level $H=\text{const}$ are proportional with a constant factor.

**Theorem 13.13.** *Let a (local) one-parameter group of transformations $\Phi_t$ correspond to a Hamiltonian $H(x,p)$ such that $H(x,\lambda p)=\lambda H(x,p)$ for $\lambda>0$. Then all the transformations $\Phi_t$ preserve the form $pdx=p_i\,dx^i$.*

**Proof.** Consider the transformations $\Phi_t$ for $t$ close to some $t_0$. Without loss of generality we may assume that $t_0=0$. Then we have

$$p_i\to p_i'=p_i-\frac{\partial H}{\partial x^i}t,\quad x^i\to x'^i=x^i+\frac{\partial H}{\partial p_i}t,$$
$$dx^i\to dx'^i=dx^i+t\,d\Big(\frac{\partial H}{\partial p_i}\Big).$$

All these equalities hold within $O(t^2)$, since

$$\Phi_t(p_i)=p_i'=p_i+\dot p_it+\cdots,$$
$$\Phi_t(x^i)=x'^i=x^i+\dot x^it+\cdots.$$

The form $pdx$ transforms into

$$\begin{aligned}p_i'\,dx'^i&=\Big(p_i-t\frac{\partial H}{\partial x^i}\Big)\Big(dx^i+t\,d\Big(\frac{\partial H}{\partial p_i}\Big)\Big)+O(t^2)\\&=p_i\,dx^i+t\Big[-\frac{\partial H}{\partial x^i}\,dx^i+p_i\,d\Big(\frac{\partial H}{\partial p_i}\Big)\Big]+O(t^2).\end{aligned}$$

Since $p_i\,d(\frac{\partial H}{\partial p_i})=d(p_i\,\frac{\partial H}{\partial p_i})-\frac{\partial H}{\partial p_i}\,dp_i$, we have

$$p_i'\,dx'^i=p_i\,dx^i+t\Big[-dH+d\Big(p_i\,\frac{\partial H}{\partial p_i}\Big)\Big]+O(t^2).$$

In view of $H(x, \lambda p) = \lambda H(x,p)$,

$$p_i \frac{\partial H}{\partial p_i} = \frac{\partial H(x, \lambda p)}{\partial \lambda} = H(x,p),$$

and we finally obtain

$$p'\,dx' = p\,dx + O(t^2).$$

This implies that the Lie derivative $\frac{d}{dt}(p\,dx)$ vanishes, i.e., $p\,dx$ does not change with time. Hence the theorem. □

A surface $\Gamma$ in the phase space is called a *conic Lagrangian surface* if the restriction of the form $p\,dx$ to the surface $\Gamma$ is identically equal to zero.

Theorem 13.13 implies the following result.

**Corollary 13.5.** *Hamiltonian systems such that $H(x, \lambda p) = \lambda H(x,p)$ for $\lambda > 0$ preserve the class of conic Lagrangian surfaces.*

EXAMPLE. The Lagrangian surface $\Gamma_{x_0}$ ($x = x_0$ with arbitrary momenta) is conic, since $p\,dx = 0$ on it. On the contrary, the surface $\Gamma_{p_0}$ ($p = p_0 \neq 0$ with arbitrary coordinates) is not conic ($p\,dx \neq 0$).

For any Hamiltonian which is first-order homogeneous in momentum, the surface $\Gamma_{x_0}$ is associated with an important bundle of trajectories emanating from the point $x_0$. Denote by $S^{n-1}_{x_0}$ the surface of dimension $n-1$ obtained as the intersection of $\Gamma_{x_0}$ with an energy level surface $H = E_0$. Consider all the trajectories starting from the points of $S^{n-1}_{x_0}$. At any time $t > 0$ we obtain the surface $S^{n-1}_{x_0}(t)$. The projection of $S^{n-1}_{x_0}(t)$ into the $x$-space is called the *wave front* at time $t$.

This is a surface of dimension $n-1$ in $x$-space depending on time $t$. Obviously, for $t = 0$ the projection of the surface $S^{n-1}_{x_0}(0)$ into the $x$-space is a single point $x_0$.

For wave front surfaces the following *Huygens principle* holds:

*To obtain the wave front surface for $t_1 > t_0 > 0$ from that at $t_0$, one must consider all the wave fronts for time $t_1 - t_0$ centered at the points of the wave front surface at time $t_0$. Then their envelope will be the wave front for time $t_1$.*

**Lemma 13.17.** *The projection of an $n$-dimensional conic Lagrangian surface $\Gamma^n$ to the $x$-space has dimension at most $n-1$.*

**Proof.** Since the $x$-components $\xi^i$ of tangent vectors to $\Gamma$ are subject to the linear relation $p_i\xi^i = 0$, the dimension of the tangent space to the projection is no greater than $n-1$. Hence the lemma. □

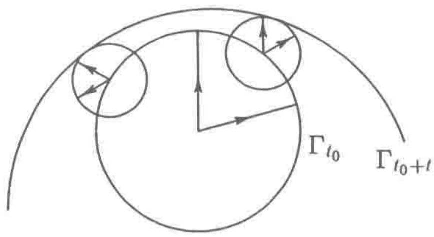


**Figure 13.2.** Wave front and the Huygens principle.

We see that a conic Lagrangian surface $\Gamma^n$ cannot be specified as a graph $p_i = f_i(x)$ even locally, and no function $S(x)$ can be assigned to it.

Consider the intersection of $\Gamma^n$ with the level surface $H(x,p) = E_0$ and denote it by $S^{n-1}$. According to Corollary 13.4, the bundle of trajectories starting from $S^{n-1}$ yields a Lagrangian surface $\hat{\Gamma}^n = \bigcup_{-\infty<t<\infty} S^{n-1}(t)$ on the level surface $H(x,p) = E_0$, which consists entirely of trajectories. Unlike $\Gamma^n$, the surface $\hat{\Gamma}^n$ may have (locally) the form $p_i = \frac{\partial S_0(x)}{\partial x^i}$. The function $S_0(x)$ (truncated action) satisfies the Hamilton–Jacobi equation

$$E_0 = H\Big(x, \frac{\partial S_0(x)}{\partial x}\Big)$$

(recall that the Hamiltonian does not depend on time).

Since the surface $\Gamma^n$ is conic, the form $p\,dx$ vanishes on $S^{n-1}$. Hence this form vanishes also on each surface $S^{n-1}(t)$ (by Theorem 13.13). Therefore, the function

$$S_0(P) = \int_{P_0}^{P} p\,dx,$$

specifying the Lagrangian surface $\hat{\Gamma}^n$, is constant on $S^{n-1}(t)$ for any $t$. Moreover, $\hat{\Gamma}^n$ lies on the energy level $H = E_0$. Therefore, $E_0 = H\Big(x, \frac{\partial S_0}{\partial x}\Big)$, and the level surfaces $S_0(x) = \text{const}$ are precisely the projections of the surfaces $S^{n-1}(t)$ into the $x$-space. Thus we have proved the following theorem.

**Theorem 13.14.** *The wave front surfaces are precisely the level surfaces* $S_0(x) = \text{const}$ *of the truncated action* $S_0$.

**13.2.4. The "action-angle" variables.** In the examples of integrable problems given above (geodesics on surfaces of revolution, the two-particle problem in a field with translation-invariant potential) the solution was obtained with the aid of additional (apart from the Hamiltonian) integrals of motion. By restricting the system to their invariant level surfaces we reduced the number of variables and hence the complexity of the system.

Now we present the Liouville integration method, which provides a rather general procedure for integrating Hamiltonian systems.

Let $M^{2n}$ be a symplectic manifold of dimension $2n$, and let $H$ be a smooth function on $M^n$ specifying a Hamiltonian system. By formula (13.6) the variation of any smooth function $f$ on the manifold $M^{2n}$ along the trajectories of the system is governed by the Liouville equation

$$\dot{f} = \{f, H\},$$

where the Poisson bracket corresponds to the symplectic structure.

A Hamiltonian system with Hamiltonian $H$ on $M^n$ is said to be *totally integrable* or *Liouville integrable* if, apart from the Hamiltonian $H = H_1$, there are $n-1$ integrals of motion (first integrals) $H_2, \ldots, H_n$ satisfying the following conditions.

1) These integrals are functionally independent almost everywhere, i.e., the Jacobian of the mapping

$$F\colon M^{2n} \to \mathbb{R}^n, \quad F(y) = (H_1(y), \ldots, H_n(y))$$

has maximal rank $n$ outside a set of points of zero measure (the mapping $F$ is called the *momentum map*).

2) The first integrals $H_1, \ldots, H_n$ are in involution, i.e., they commute pairwise in terms of Poisson brackets,

$$\{H_i, H_j\} = 0, \quad i, j = 1, \ldots, n.$$

Since we have a symplectic structure, condition 1) implies that the vectors $\nabla H_1, \ldots, \nabla H_n$ are linearly independent almost everywhere. Now we recall that by Theorem 13.7

$$\nabla\{H_i, H_j\} = -[\nabla H_i, \nabla H_j], \quad i, j = 1, \ldots, n.$$

Condition 2) now implies that if we consider $n$ Hamiltonian systems with Hamiltonians $H_1, \ldots, H_n$, then translations along the trajectories of these systems are pairwise commuting, since

$$[\nabla H_i, \nabla H_j] = 0, \quad i, j = 1, \ldots, n.$$

The image of the momentum map has positive measure, since it has maximal rank almost everywhere. By the Sard theorem, the set of critical values of the momentum map has zero measure. Therefore, for almost all values $c = (c_1, \ldots, c_n)$ of the momentum map, the inverse image $F^{-1}(c)$ is a smooth submanifold in $M^{2n}$.

Consider now the setting where these submanifolds are compact. This holds, e.g., for Hamiltonian systems with Hamiltonian

$$H(x, p) = \frac{1}{2} g^{ij}(x) p_i p_j + U(x)$$

on $T^*N$ if the potential is bounded and the manifold $N$ is compact. In this case any level surface $H = \text{const}$ is compact.

**Lemma 13.18.** *Let $X^n$ be a connected compact submanifold in $M^{2n}$ lying in the intersection of level surfaces $H_1 = c_1, \dots, H_n = c_n$ of the first integrals $H_1, \dots, H_n$, and these first integrals are functionally independent at all the points of $X^n$. Then $X^n$ is an $n$-dimensional torus.*

**Proof.** Since $X^n$ is specified (locally) by $n$ equations $H_1 = c_1, \dots, H_n = c_n$ in a $2n$-dimensional manifold and these equations are independent, by the implicit function theorem $X^n$ is of dimension $n$.

On $X^n$ we have the action of the group $\mathbb{R}^n$ generated by translations along the trajectories of Hamiltonian flows with Hamiltonians $H_1, \dots, H_n$. The orbit of any point has dimension $n$ and is a submanifold in $X^n$, since at any point of $X^n$ the vectors $\nabla H_1, \dots, \nabla H_n$ are linearly independent. Therefore, this orbit coincides with $X^n$.

All the translations in $\mathbb{R}^n$ which map a point $x \in X^n$ into itself form a subgroup $\Gamma \subset \mathbb{R}^n$, which is, obviously, a lattice and $X^n$ is the quotient space $X^n = \mathbb{R}^n/\Gamma$. Since the manifold $X^n$ is compact, this lattice has maximal rank (is isomorphic to $\mathbb{Z}^n$), and we have

$$X^n = \mathbb{R}^n/\Gamma \approx \mathbb{R}^n/\mathbb{Z}^n = T^n.$$

Hence the lemma. □

Consider now a neighborhood of such a torus $X^n$, e.g., the inverse image $U$ of an open set $V \subset \mathbb{R}^n$ under the momentum map: $U = F^{-1}(V)$.

**Theorem 13.15.** *Suppose that the domain $U$ in $M^{2n}$ is filled out by $n$-dimensional invariant tori $X(c)$ of the form $H_1 = c_1, \dots, H_n = c_n$, $c = (c_1, \dots, c_n)$, and the first integrals $H_1, \dots, H_n$ are functionally independent at the points of $U$.*

*Then for any torus $X(c')$ there is a neighborhood $U$ filled out by tori with values $F = c$ close to $c'$, in which one can introduce coordinates $I_1, \dots, I_n$, $\varphi_1, \dots, \varphi_n$ such that:*

1) $I_k = I_k(c_1, \dots, c_n)$, $k = 1, \dots, n$.

2) *The coordinates $\varphi_1, \dots, \varphi_n$ are defined modulo $2\pi$, and for any fixed $I = (I_1, \dots, I_n)$ the points $y = (I_1, \dots, I_n, \varphi_1, \dots, \varphi_n)$ fill out the $n$-dimensional torus $X(c), I = I(c)$ (i.e., the variables $\varphi_k$, $k = 1, \dots, n$, are cyclic coordinates on the tori $X(c)$).*

3) *In these coordinates the symplectic form becomes*

$$\Omega = \sum_{k=1}^{n} dI_k \wedge d\varphi_k. \tag{13.10}$$

4) *The Hamilton equations for the Hamiltonian* $H = H_1(I_1, \dots, I_n)$ *in these coordinates have the form*

$$\dot{I}_k = 0, \quad \dot{\varphi}_k = \omega_k(I_1, \dots, I_k), \quad k = 1, \dots, n. \tag{13.11}$$

The variables $I_k, \varphi_l$, $k, l = 1, \dots, n$, satisfying the conditions of this theorem are called the *"action-angle" variables.* Equations (13.11) follow from equality (13.10) and the fact that the Hamiltonian depends only on the action variables $I_1, \dots, I_n$ and is independent of the angular variables $\varphi_1, \dots, \varphi_n$. We see that the motion on invariant tori is linear in these variables.

Construction of "action-angle" variables enables one to integrate the problem, since equations (13.11) can be integrated explicitly: for initial conditions $I = I'$, $\varphi = \varphi'$ at $t = 0$, the solution is

$$I_k(t) = I'_k, \quad \varphi_l(t) = \varphi'_l + \omega_l(I'_1, \dots, I'_n)t, \quad k, l = 1, \dots, n.$$

Note that equality (13.10) implies that the invariant tori $X(c)$ are Lagrangian.

**Proof of Theorem 13.15.** Take an $n$-dimensional surface $Y$ that intersects each torus $X(c)$ close to $X(c')$ at a single point $x(c)$. We may assume that these tori fill out a domain $U = F^{-1}(V)$, where $V = F(Y) \subset \mathbb{R}^n$ is the spherical domain of values of the momentum map.

Each point $y \in X(c)$ is obtained from the point $x(c)$ by a combination of translations for times $t_k$ along the Hamiltonian flows with Hamiltonians $H_k$ (their ordering is immaterial because these flows commute). Take the tuple $(t_1, \dots, t_k)$ for the coordinates of $y$.

Thus we obtain the coordinates $(t_1, \dots, t_n, H_1, \dots, H_n)$ satisfying the relations

$$\{t_i, H_j\} = \delta_{ij}, \quad \{H_i, H_j\} = 0, \quad i, j = 1, \dots, n$$

(the first equality is a consequence of the fact that the $t_i$ are the times along the trajectories of the flows). Then the symplectic form is

$$\Omega = \sum_i dH_i \wedge dt_i + \sum c_{ij}\, dH_i \wedge dH_j.$$

On each torus, take a basis $e_1, \dots, e_n$ in the lattice of periods $\Gamma$. This may be done continuously by shrinking the domain $U$ if necessary. Now in each torus we make a linear change of coordinates

$$(t_1, \dots, t_n) \to (\varphi_1, \dots, \varphi_n)$$

such that the basis vectors of the lattice take a simple form

$$e_i = (0, \dots, 0, 2\pi, 0, \dots, 0), \quad i = 1, \dots, n$$

(with $2\pi$ at the $i$th place and zeros otherwise). These are angular coordinates (up to a translation). When the $i$th coordinate varies from 0 to $2\pi$ (with the rest fixed), the point on the torus runs through a closed contour $a_i$.

Since the change $t \to \varphi$ depends on the values $H_1, \dots, H_n$, the symplectic structure in the new coordinates becomes

$$\Omega = \sum_{k,l} a_{kl}(H, \varphi)\, dH_k \wedge dH_l + \sum_{k,l} b_{kl}(H)\, dH_k \wedge d\varphi_l. \tag{13.12}$$

This implies that all the invariant tori $H_1 = c_1, \dots, H_n = c_n$ are Lagrangian, and shrinking the neighborhood $U$ homotopically to the torus (by shrinking the domain $V$ into the point $c'$), we deduce from the homotopic invariance of the de Rham cohomology (see 9.3.2) that $[\Omega] = 0$ in $H^2(U; \mathbb{R})$, since $[\Omega] = 0$ in $H^*(X(c'); \mathbb{R})$. Therefore, the form $\Omega$ is exact in the domain $U$, and we can choose the form $\alpha = d^{-1}\Omega$ to be well defined in the entire domain $U$.

Define the action variables by the formula

$$I_j = \frac{1}{2\pi} \int_{a_j} \alpha, \quad j = 1, \dots, n, \tag{13.13}$$

where the basis contours $a_j$ are taken on the torus $X(H_1, \dots, H_n)$. The functions $I_1, \dots, I_n$ do not depend on $\varphi_1, \dots, \varphi_n$. This is shown as follows.

If we take any other contour $a'_j$, which together with $a_j$ encloses a surface $T$ on the torus, then by the Stokes theorem we obtain

$$\int_{a_j} \alpha - \int_{a'_j} \alpha = \int_T d\alpha = \int_T \Omega = 0,$$

since the restriction of the form $\Omega$ to the torus is equal to zero. This implies that the integrals of $\alpha$ over all contours homotopic to $a_j$ are equal. In fact the value of the integral depends only on the cohomology class realized by the contour $a_j$.

Formula (13.12) implies that

$$\alpha = \sum_l \Big( \int \sum_k b_{kl}\, dH_k \Big)\, d\varphi_l + \sum_k f_k\, dH_k + df_0,$$

and the result of integration over the cycles $a_j$ is determined solely by the first term,

$$dI_l = \sum_k b_{kl}\, dH_k \tag{13.14}$$

(the integral formulas determine the coordinates $I_k$ up to constants). Hence

$$\frac{\partial I_l}{\partial H_k} = b_{kl}(H_1, \dots, H_n),$$

and since the form $\Omega$ is nondegenerate, the matrix $b_{kl}$ is invertible. Therefore, by the inverse function theorem there exists locally a change of coordinates

$$H_1, \ldots, H_n \to I_1, \ldots, I_n, \quad I_k = J_k(H_1, \ldots, H_n),$$

which does not affect the other coordinates $\varphi_1, \ldots, \varphi_n$.

The form $\Omega$ in the new variables becomes

$$\Omega = \sum_{k,l} a'_{kl}(I, \varphi)\, dI_k \wedge dI_l + \sum_{k,l} b'_{kl}\, dI_k \wedge d\varphi_l.$$

Applying to $\Omega$ the above argument, we obtain from (13.13) an analog of (13.14):

$$dI_l = \sum_l b'_{kl}\, dI_k.$$

Since the variables $I_1, \ldots, I_n$ are independent, we have $b'_{kl} = \delta_{kl}$, and $\Omega$ takes the almost canonical form

$$\Omega = \sum_{k,l} a'_{kl}\, dI_k \wedge dI_l + \sum_k dI_k \wedge d\varphi_k.$$

The form $\Omega$ is closed, which means that the coefficients $a'_{kl}$ do not depend on the variables $\varphi_i$ and the form $\beta = \sum a'_{kl}\, dI_k \wedge dI_l$ is closed. We have locally

$$\beta = d\Big(\sum_k g_k(I)\, dI_k\Big).$$

It remains to shift the angular variables,

$$\varphi_k \to \varphi_k + g_k(I),$$

to reduce $\Omega$ to the canonical form

$$\Omega = \sum_k dI_k \wedge d\varphi_k.$$

Thus the proof of Theorem 13.15 is completed. □

## 13.3. Local minimality condition

**13.3.1. The second-variation formula and the Jacobi operator.** The Euler–Lagrange equations

$$\frac{\delta S}{\delta x^i} = \frac{\partial L}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = 0, \quad i = 1, \ldots, n,$$

for a functional

$$S[\gamma] = \int_\gamma L(x, \dot{x})\, dt$$

are only a necessary condition for $S[\gamma]$ to have a minimum along some curve $\gamma$ among, e.g., all curves joining some points $P$ and $Q$.

For functions $f(x^1, \dots, x^n)$ of many variables, a sufficient condition for a local minimum at a given critical point $P$ is that the Hessian, i.e., the quadratic form $d^2 f$, is positive definite at $P$. Hence, to find a local minimum of $S[\gamma]$, where the curve $\gamma$ satisfies the Euler–Lagrange equations, we must find the second variation $\delta^2 S_\gamma(\eta, \xi)$ of the functional $S$:

$$\frac{\partial^2}{\partial\lambda\partial\mu} S[\gamma + \lambda\xi + \mu\eta]\Big|_{\lambda=0,\mu=0} = \delta^2 S_\gamma(\xi, \eta) = \delta^2 S_\gamma(\eta, \xi),$$

where $\eta$, $\xi$ are vector fields defined on the curve $\gamma(t)$ and vanishing at the points $\gamma(a) = P$ and $\gamma(b) = Q$.

**Lemma 13.19.** *If the curve $\gamma$: $\{x^i = x^i(t)\}$ satisfies the Euler–Lagrange equations, then the following* formula for the second variation *holds:*

$$\delta^2 S_\gamma(\xi, \eta) = \frac{\partial^2 S[\gamma + \lambda\xi + \mu\eta]}{\partial\lambda\partial\mu}\Big|_{\lambda=0,\mu=0} = -\int_a^b (J_{ij}\xi^j)\eta^i\,dt,$$

*where*

$$J_{ij}\xi^j = \frac{d}{dt}\Big(\frac{\partial^2 L}{\partial\dot{x}^i\partial\dot{x}^j}\,\dot{\xi}^j + \frac{\partial^2 L}{\partial\dot{x}^i\partial x^j}\,\xi^j\Big) - \frac{\partial^2 L}{\partial x^i\partial\dot{x}^j}\,\dot{\xi}^j - \frac{\partial^2 L}{\partial x^i\partial x^j}\,\xi^j. \tag{13.15}$$

**Proof.** Using the formula for the first variation (see Section 12.2), we obtain

$$\begin{aligned}\frac{\partial^2 S[\gamma + \lambda\xi + \mu\eta]}{\partial\lambda\partial\mu}\Big|_{\lambda=0,\mu=0} &= \Big[\frac{\partial}{\partial\lambda}\Big(\frac{\partial}{\partial\mu} S[\gamma + \lambda\xi + \mu\eta]\Big)\Big|_{\mu=0}\Big]\Big|_{\lambda=0} \\ &= \frac{\partial}{\partial\lambda}\int_a^b \Big(\frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial\dot{x}^i}\Big)\eta^i\,dt\Big|_{\lambda=0} \\ &= \int_a^b \Big(\frac{\partial^2 L}{\partial x^i\partial x^j}\,\xi^j + \frac{\partial^2 L}{\partial x^i\partial\dot{x}^j}\,\dot{\xi}^j - \frac{d}{dt}\Big(\frac{\partial^2 L}{\partial\dot{x}^i\partial x^j}\,\xi^j + \frac{\partial^2 L}{\partial\dot{x}^i\partial\dot{x}^j}\,\dot{\xi}^j\Big)\Big)\eta^i\,dt.\end{aligned}$$

Hence the lemma. □

A linear operator $J$ as in (13.15), acting on vector fields $\xi(t)$ defined on the curve $\gamma$, is called the *Jacobi operator.*

The Jacobi operator $J$ is defined as a vector linear differential operator acting of vector-functions of one variable $t$ by the formula

$$J_{ij}\xi^j = A_{ij}(t)\frac{d^2\xi^j}{dt^2} + B_{ij}(t)\frac{d\xi^j}{dt} + C_{ij}(t)\xi^j(t),$$

provided some local coordinates along the curve are chosen. This operator determines the scalar product

$$\langle J\xi, \eta\rangle = \int J_{ij}\xi^j(t)\eta^i(t)\,dt$$

on a space of vector-functions $(\xi^i(t))$, $(\eta^j(t))$. The definition of the scalar product requires also specification of boundary conditions. The most widely used boundary conditions are as follows.

1. *Dirichlet condition.* In this case the vector-functions $\xi^i(t)$, $\eta^i(t)$ are assumed to vanish at the endpoints $t = t_0$ and $t = t_1$. These values correspond to two points on the extremal $\gamma$,

$$\gamma(t_0) = P, \quad \gamma(t_1) = Q.$$

Thus we explore our variational problem in the class of curves starting at the point $P$ and ending at the point $Q$.

2. *Periodic problem.* Here the coefficients of the operator $J$ are periodic with some period $T$ in the variable $t \in \mathbb{R}$. However, we study the operator $J$ not only on periodic vector-functions. Such problems are connected with investigation of a neighborhood of periodic extremals. We will not discuss them in detail here.

3. *General separated boundary conditions.* The vector fields $\xi(t)$, $\eta(t)$ for $t = t_0$ and $t = t_1$ must satisfy the conditions

$$\begin{aligned} t = t_0: &\quad u_0\xi + v_0\dot{\xi} = 0, \\ t = t_1: &\quad u_1\xi + v_1\dot{\xi} = 0, \end{aligned}$$

where $u_0, v_0, u_1, v_1$ are $n \times n$ matrices, and the curve $\gamma$ belongs to an $n$-dimensional manifold $N$, $i, j = 1, \dots, n$.

The Lagrangian $L$ is said to be nondegenerate if at any point $x \in N$

$$\det\left(\frac{\partial^2 L}{\partial \dot{x}^i \partial \dot{x}^j}\right)_x \neq 0.$$

Recall that the Lagrangians for which the Legendre transform $p = \frac{\partial L}{\partial \dot{x}}$ is globally invertible,

$$(x, \dot{x}) \leftrightarrow (x, p),$$

were called strongly nondegenerate.

Separated boundary conditions are specified as follows. It is convenient to treat extremal curves as lying in the space

$$M \times I = T^*N \times I,$$

where $I = [t_0, t_1]$ is a chosen time interval, $N$ is the configuration space of the problem with Lagrangian $L(x, \dot{x}, t)$, and $T^*(N)$ is the corresponding phase space. We assume the Lagrangian to be strongly nondegenerate, and so the Euler–Lagrange equations are equivalent to the Hamilton equations. Then for two given $n$-dimensional Lagrangian submanifolds,

$$\begin{aligned} L_0 &\subset T^*(N) \times t_0, \\ L_1 &\subset T^*(N) \times t_1, \end{aligned}$$

the extremal is required to start in $L_0$ and end in $L_1$:

$$\gamma(t_0) = P \in L_0, \quad \gamma(t_1) = Q \in L_1.$$

The following problems provide the simplest examples of the above settings.

1. The *Dirichlet problem* where both manifolds $L_0$ and $L_1$ are momentum $p$-spaces at the points $P = \gamma(t_0)$ and $Q = \gamma(t_1)$. This is the problem with fixed endpoints.

2. The *Neumann problem* where the manifolds $L_0$ and $L_1$ are specified as $(n-1)$-dimensional surfaces in the configuration $x$-space $N$ and the extremal curves are required to be normal to them. The surface $L_0$ in coordinates $(x^1, \dots, x^n, p_1, \dots, p_n)$ is locally, in a neighborhood of the point $P = \gamma(t_0)$, representable as

$$L_0 = \{x^n = 0, p_1 = \cdots = p_{n-1} = 0\}.$$

Such local coordinates for the Neumann problem may be chosen in $N$ and $T^*(N)$, and similarly for $L_1$.

Thus a problem with separated boundary conditions is a generalization of the Dirichlet and Neumann problems.

For nondegenerate Lagrangians, the Jacobi equation

$$J\xi = 0 \quad \text{or} \quad J_{ij}\xi^j = 0, \quad i = 1, \dots, n,$$

has a nondegenerate leading term and is equivalent to a linear Hamilton equation in the variables

$$p_i = A_{ij}(t)\dot{\xi}^j, \quad \tilde{x}^i = \xi^i = x^i,$$

where the Hamiltonian is quadratic in $(x, p)$ and depends on $t$,

$$H(x, p, t) = \frac{1}{2} A^{ij} p_i p_j + \tilde{B}^i_j p_i x^j + \tilde{C}_{ij} x^i x^j,$$

with $A^{ij} A_{jk}(t) = \delta^i_k$. All the other coefficients can be easily obtained from the formulas

$$H = \dot{x}^i \frac{\partial L}{\partial \dot{x}^i} - L, \qquad p_i = \frac{\partial L}{\partial \dot{x}^i} = A_{ij}\dot{x}^j.$$

A Lagrangian is said to be *regular* if the quadratic form

$$\frac{\partial^2 L}{\partial \dot{x}^i \partial \dot{x}^j} \eta^i \eta^j$$

is positive for $\eta \neq 0$. The following fact is known from the theory of differential equations. If the Lagrangian is regular for all $t \in \mathbb{R}$, then for any $t \in \mathbb{R}$ there exists $\varepsilon > 0$ such that the Jacobi quadratic form

$$\langle J\eta, \eta \rangle = \int_{t'}^{t''} J_{ij} \eta^i \eta^j \, dt$$

is negative ($\langle J\eta, \eta \rangle < 0$) on the space of nonzero vector-functions $\eta(t)$ such that

$$\eta(t') = \eta(t'') = 0, \quad |t' - t''| \leq \varepsilon, \quad t' \leq t \leq t''.$$

Hence we arrive at the following conclusion.

*Suppose $N$ is a compact manifold and*

$$\frac{\partial^2 L}{\partial \dot{x}^i \partial \dot{x}^j} \eta^i \eta^j > 0 \quad \text{for} \quad \eta \neq 0.$$

*Then for any extremal curve $\gamma(t)$ (i.e., any solution to the Euler–Lagrange equations) and any pair of its points sufficiently close to each other, $P = \gamma(t_0)$, $Q = \gamma(t_1)$, $|t_0 - t_1| < \varepsilon$, the extremal is an isolated local minimum among all smooth curves joining the points $P$ and $Q$.*

Now we turn to the theory of geodesics.

As an example we point out that for the Lagrangian $L = \sqrt{g_{ij}(x)\dot{x}^i\dot{x}^j}$ (i.e., for the length functional $S = l$), the corresponding matrix is such that

$$\det\Big(\frac{\partial^2 L}{\partial \dot{x}^i \partial \dot{x}^j}\Big) = 0.$$

Hence the corresponding Jacobi form is not strictly positive. This was to be expected: our functional remains unchanged under reparametrizations, which implies that there are zero eigenvalues of the Jacobi quadratic form.

For geodesics, it is expedient to consider the action

$$S = \int_a^b \frac{1}{2}\, g_{ij}\dot{x}^i\dot{x}^j\, dt.$$

**Theorem 13.16.** *The bilinear form*

$$\frac{\partial^2 S[\gamma + \lambda\xi + \mu\eta]}{\partial\lambda\partial\mu}\Big|_{\lambda=0,\mu=0} = \delta^2 S_\gamma(\xi, \eta)$$

*for the action $S = \int_a^b \frac{1}{2} g_{ij}\dot{x}^i\dot{x}^j\, dt$, any geodesic $\gamma$: $x^i = x^i(t)$, and smooth vector fields $\xi(t), \eta(t)$ along the geodesic that vanish at its endpoints, can be written as*

$$\delta^2 S_\gamma(\xi, \eta) = -\int_a^b \langle \nabla_{\dot{x}}^2 \xi + R(\dot{x}, \xi)\dot{x}, \eta\rangle\, dt,$$

*or*

$$\delta^2 S_\gamma(\xi, \eta) = -\int_a^b \langle J\xi, \eta\rangle\, dt, \tag{13.16}$$

*where $t$ is the natural parameter along the geodesic $\gamma$ and*

$$(J\xi)^i = \nabla_{\dot{x}}^2 \xi^i + \dot{x}^j\dot{x}^l\xi^k R^i_{jkl}$$

*with $R^i_{jkl}$ being the curvature tensor.*

**Proof.** For the Lagrangian $L = \frac{1}{2}\, g_{ij}\dot{x}^i\dot{x}^j$, the first variation is given by the formula (see Section 12.2)

$$\frac{\partial}{\partial\lambda} S[\gamma + \lambda\xi]\Big|_{\lambda=0} = -\int_a^b (\ddot{x}^k + \Gamma^k_{ij}\dot{x}^i\dot{x}^j) g_{kl}\xi^l\, dt = -\int_a^b \langle \nabla_{\dot{x}}\dot{x}, \xi\rangle\, dt.$$

Hence

$$\frac{\partial^2}{\partial\lambda\partial\mu} S[\gamma+\lambda\xi+\mu\eta]\Big|_{\lambda=0,\mu=0}$$

$$= -\frac{\partial}{\partial\lambda}\int_a^b \langle\nabla_{\dot x}\dot x,\eta\rangle\,dt\Big|_{\lambda=0} = -\int_a^b \{\langle\nabla_\xi\nabla_{\dot x}\dot x,\eta\rangle + \langle\nabla_{\dot x}\dot x,\nabla_\xi\eta\rangle\}\,dt.$$

The second term in the integrand vanishes on $\gamma$ in view of the equation for geodesics: $\nabla_{\dot x}\dot x = 0$. We transform the first term. We have

$$\nabla_\xi\nabla_{\dot x} - \nabla_{\dot x}\nabla_\xi = R(\dot x,\xi), \quad \nabla_\xi\dot x = \nabla_{\dot x}\xi,$$

by symmetry of the connection. Finally, we obtain

$$\nabla_\xi\nabla_{\dot x}\dot x = \nabla^2_{\dot x}\xi + R(\dot x,\xi)\dot x,$$

which implies the theorem. □

EXAMPLE. Consider the two-dimensional case. Near the geodesic $\gamma(t)$ we introduce a special system of coordinates $(x, y)$, $x = x^1$, $y = x^2$ (semi-geodesic coordinates; see 10.3.2) in which:

a) the line $(x, 0)$ is the geodesic $\gamma$ itself, and $x$ is the natural parameter;

b) the coordinate $y$ is orthogonal to the geodesic and $g_{ij}(x,0) = \delta_{ij}$.

In this case the bilinear form $\delta^2 S_\gamma(\xi,\eta)$ for the pair of fields $\xi(t)$, $\eta(t)$ normal to $\gamma(t)$ can be written as

$$\delta^2 S_\gamma(\xi,\eta) = -\int_a^b \Big(\frac{d^2}{dt^2}\xi^i + K(t)\xi^i\Big)\eta_i\,dt,$$

where $K$ is the Gaussian curvature. Note that $\delta^2 S_\gamma(\xi,\eta) \equiv 0$ if at least one of these fields is proportional to the tangent vector field $\dot x(t)$ with a constant factor. However, this variation does not satisfy the boundary conditions because it does not vanish at the endpoints.

It is not hard to show that for piecewise smooth vector fields $\xi(t)$ the bilinear form $\delta^2 S_\gamma(\xi,\eta)$ is

$$\delta^2 S_\gamma(\xi,\eta) = -\sum_{P_i}\langle\eta,\Delta_{P_i}(\nabla_{\dot x}\xi)\rangle - \int_a^b \langle J\xi,\eta\rangle\,dt,$$

where the summation extends over all the discontinuity points $P_i$ of $\nabla_{\dot x}\xi$, and $\Delta_P(\nabla_{\dot x}\xi)$ is the jump of the covariant derivative at the discontinuity point $P$.

The proof of this formula follows the lines of that of Theorem 13.16 taking into account the discontinuity points when integrating by parts.

**13.3.2. Conjugate points.** Now, let us find out under which conditions the bilinear form $\delta^2 S_\gamma(\xi, \eta)$ is nondegenerate. For simplicity, we will consider only geodesics, although this is not crucial.

A vector field $\xi$ along an extremal $\gamma$ going from $P$ to $Q$ is said to be a *Jacobi field* if it satisfies the Jacobi equation $J\xi = 0$ and vanishes at the endpoints $P$ and $Q$.

For $S = \int \langle \dot{x}, \dot{x} \rangle \, dt$ we have

$$J\xi^i = \nabla^2_{\dot{x}} \xi^i + \dot{x}^j \dot{x}^l \xi^k R^i_{jkl} = 0, \quad i = 1, \dots, n.$$

Points $P$ and $Q$ are said to be *conjugate* along the extremal—the geodesic $\gamma$ joining $P$ and $Q$—if there is a nonzero Jacobi field $\xi$ along $\gamma$.

**Lemma 13.20.** *The bilinear form $\delta^2 S_\gamma(\xi, \eta)$ is nondegenerate on the space of smooth fields vanishing at the endpoints if and only if the endpoints $P$ and $Q$ of the extremal $\gamma$ are not conjugate along $\gamma$.*

**Proof.** Actually this lemma says that any bilinear form on a space with nondegenerate scalar product is specified by the Jacobi operator, and the "zeros" of this operator (the vectors in its kernel), if any, make the bilinear form degenerate. Thus this lemma restates a well-known fact of linear algebra. □

This lemma remains valid also on the space of piecewise smooth vector fields.

**Theorem 13.17.** *For a positive Riemannian metric, if a geodesic $\gamma$ joining some points $P$ and $Q$ contains a pair of conjugate points $P'$, $Q'$, then this geodesic $\gamma$ is not minimal.*

**Proof.** We may assume that the points $P$ and $Q$ are not conjugate along $\gamma$ (otherwise, we can move inside the curve and take the resulting geodesic $\gamma' \subset \gamma$). We will consider the form extended to all piecewise smooth curves, which is also nondegenerate. Then the bilinear form $\delta^2 S_\gamma(\xi, \eta)$ is nondegenerate. Therefore, the quadratic form $\delta^2 S_\gamma(\xi, \xi)$ for a minimal geodesic must be nonnegative, $\delta^2 S_\gamma(\xi, \xi) \geq 0$. We will show that this requirement fails in the presence of conjugate points.

Let $\xi'$ be the Jacobi field corresponding to the points $P'$, $Q'$. First, we construct a piecewise smooth field $\xi$ between $P$ and $Q$ which is equal to $\xi'$ between$P'$ and $Q'$ and zero outside the interval $P'Q'$ (see Figure 13.3). Then we have

$$\delta^2 S_\gamma(\xi, \xi) = 0.$$

However, the form $\delta^2 S_\gamma$ is nondegenerate. Therefore, $\delta^2 S_\gamma(\xi, \xi)$ takes also negative values. Hence the theorem. □

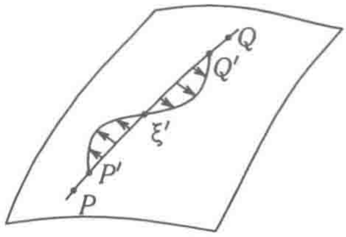


**Figure 13.3.** Conjugate points and the field $\xi$.

**Theorem 13.18.** *For a sufficiently small length interval, the geodesics of a Riemannian metric afford the global minimum of the action functional $S[\gamma]$ among all smooth curves joining the same points. Therefore, any sufficiently short geodesic locally realizes the shortest path between the points.*

**Proof.** A sufficient condition for local minimality of a geodesic $\gamma(t)$ is that the quadratic form $\delta^2 S_\gamma(\xi, \xi)$ is positive for all vector fields $\xi$ vanishing at the endpoints. By formula (13.16) we obtain

$$\begin{aligned}\delta^2 S_\gamma(\xi, \xi) &= -\int_a^b [\langle \nabla_{\dot{x}}^2 \xi, \xi\rangle + \langle R(\dot{x}, \xi)\dot{x}, \xi\rangle]\, dt \\ &= \int_a^b [\langle \nabla_{\dot{x}}\xi, \nabla_{\dot{x}}\xi\rangle - \langle R(\dot{x}, \xi)\dot{x}, \xi\rangle]\, dt\end{aligned}$$

(integrating by parts and using the equalities $\xi(a) = \xi(b) = 0$). For a sufficiently small length interval $\Delta l$ we have the inequality

$$\left| \int_a^b \langle R(\dot{x}, \xi)\dot{x}, \xi\rangle\, dt \right| < c(\Delta l) \int_a^b \langle \nabla_{\dot{x}}\xi, \nabla_{\dot{x}}\xi\rangle\, dt,$$

where the factor $c(\Delta l)$ depends on the metric $g_{ij}$ and the length $\Delta l$ and tends to zero as $\Delta l \to 0$. Since

$$\int_a^b \langle \nabla_{\dot{x}}\xi, \nabla_{\dot{x}}\xi\rangle\, dt > 0,$$

this implies the theorem. □

## Exercises to Chapter 13

1. Let $X$ be a vector field on a configuration space. Consider the function $F_X = p_i X^i$ on the phase space. Prove that:

a) $\{F_X, F_Y\} = -F_{[X,Y]}$;

b) if a function $f = f(x)$ does not depend on momenta $p$, then $\{f, F_X\} = \partial_X f$.

2. Prove that in the Kepler problem the Lie algebra of integrals $W_i$, $M_j$ for a fixed energy $E$ is isomorphic to:

a) $\mathfrak{so}(4)$ for $E < 0$;

b) $\mathfrak{so}(1,3)$ for $E > 0$;

c) the Lie algebra of the group of motions of $\mathbb{R}^3$ for $E = 0$.

3. Let $\Omega = g_{ik}\, dy^i \wedge dy^k$ be a symplectic form, and $X = (X^k)$ a vector field. Prove that the form $i(X)(\omega) = g_{ik} X^k\, dy^i$ is closed if and only if the Lie derivative of the form $\Omega$ along the field $X$ vanishes, $L_X \Omega = 0$.

4. Let $\omega$ be a symplectic form on a manifold $M^n$. Prove that there exists an almost complex structure $J$ on $M^n$ such that the expression $\omega(u, Jv) = \langle u, v\rangle$ specifies a Riemannian metric on the manifold.

5. Let $M_x$, $M_y$, $M_z$ be the moment integrals (12.13). Prove that:

a) The function $M^2 = M_x^2 + M_y^2 + M_z^2$ is a Casimir function of the Poisson algebra generated by the moment integrals:

$$\{M^2, M_x\} = \{M^2, M_y\} = \{M^2, M_z\} = 0.$$

b) Together with the functions $p_x, p_y, p_z$, these integrals form a Lie algebra relative to the Poisson bracket, which is isomorphic to the Lie algebra of the group of motions of $\mathbb{R}^3$.

6. Prove that:

a) The formulas

$$\{x, y\} = xy, \quad \{x, z\} = xz, \quad \{x, u\} = 2yz,$$
$$\{y, z\} = 0, \quad \{y, u\} = yu, \quad \{z, u\} = zu$$

specify a Poisson structure in the space $\mathbb{R}^4$ with coordinates $x, y, z, u$.

b) The function $f(x, y, z, u) = xu - yz$ is a Casimir function of these Poisson brackets, which are quadratic in coordinates, and restriction of the Poisson structure to the surface $xu - yz = 1$ yields the invariant Poisson bracket on the Lie group

$$\mathrm{SL}(2) = \left\{ \begin{pmatrix} x & y \\ z & u \end{pmatrix} : xu - yz = 1 \right\}.$$

7. In a phase space with canonical coordinates $x^1, \ldots, x^n$, $p_1, \ldots, p_n$, we introduce new coordinates $X^1, \ldots, X^n$, $P_1, \ldots, P_n$ such that

$$p_i = \frac{\partial S(x, X)}{\partial x^i}, \quad P_i = -\frac{\partial S(x, X)}{\partial X^i},$$

$i = 1, \ldots, n$, where $S(x, X)$ is a smooth function of variables $x$, $X$ with a nondegenerate Hessian. Prove that the new coordinates are canonical. The function $S(x, X)$ is called the *generating function* of the canonical transformation $(x, p) \to (X, P)$.

8. Prove that on a manifold of negative sectional curvature geodesics contain no conjugate points.

9. a) Prove that for a manifold of positive sectional curvature $K \geq \text{const} > 0$ there exists a constant $T$ such that any geodesic of length at least $T$ contains a pair of points conjugate along this geodesic.

b) Prove that under the condition that the Ricci curvature is nonnegative, $R_{ik}\xi^i\xi^k \geq Cg_{ik}\xi^i\xi^k$, the statement a) holds for $T = (n-1)\pi/C$, where $n$ is the dimension of the manifold.

Chapter 14

# Multidimensional Variational Problems

## 14.1. Calculus of variations

**14.1.1. Introduction. Variational derivatives.** Here we use the approach to exposition of the basics of field geometry developed by theoretical physicists of the 20th century—the golden age of this science. In our opinion, this is the approach closest to geometry in its spirit and most suitable for further development of many applications both in natural sciences and geometry; it is indispensable, e.g., for understanding the field quantization. By the word "analysis" we mean approximately the same as "calculus", that is, a collection of convenient tools for working formally with analytic quantities, deriving and writing formulas, and reading and understanding them in an easiest way based on analogies with elementary classical calculus of finite-dimensional objects. The rigorous abstract language of modern function theory and functional analysis will play only a minor role in our presentation, since our purpose is the simplest possible explanation of what it is and how it is related to the finite-dimensional analysis, rather than a formal justification. We never deal with abstract infinite-dimensional manifolds. All our spaces are various function spaces in the general sense of modern viewpoint: these may be the spaces of mappings of one manifold (the domain) to another (the range), or the spaces of sections of a fiber bundle with base $X$ that plays the role of the domain. As an example of such a setting we may consider, for instance, tangent vector fields on a manifold $X$, which are sections of the tangent bundle $TX$. Anyway, locally we always deal with

vector-functions $u(x) = (y^1(x), \dots, y^m(x))$, where $x = (x^1, \dots, x^n)$ are local coordinates in the domain, and $(y^1, \dots, y^m)$ are local coordinates in the range.

An important role in finite-dimensional geometry, tensor analysis, and algebra is played by indices $(i)$ running over a finite set of values. These indices allow for construction of tensor calculus with its system of rules.

In infinite-dimensional analysis (field theory), the role of indices is played by the usual pair $(i, x)$, where $i$ runs over a finite set, $x$ over the domain (of definition), $x \in X$, and sums are replaced by integrals. In our study of the local field theory, we may initially assume that we deal simply with vector-functions defined in a domain of Euclidean space, $U \subset \mathbb{R}^m$, and taking values in another domain of Euclidean space, $V \subset \mathbb{R}^n$:

$$y^j(x^1, \dots, x^m), \quad j = 1, \dots, n.$$

This setup is appropriate for deriving local identities, relations, equations, etc. In this case we will assume that all these functions are infinitely differentiable, i.e., belong to the class $C^\infty$.

When studying variations of fields, we will first assume them to be "strongly local", i.e., belonging to $C^\infty$ and vanishing identically outside a small neighborhood of the point under consideration.

Only at a later stage, when we explicitly explore the influence of boundary conditions, we will need to admit a more general class of variations.

The problems of rigorous justification will be beyond our goals. Such problems always require completing spaces of smooth functions in some way or another. There is an extensive theory concerning these questions, but we do not plan to present it in this book.

Effects of global topological phenomena will be discussed occasionally, but only later.

Thus, for us, field theory is an infinite-dimensional analog of analysis, including tensor analysis, with the key role played in the initial presentation by the "calculus of variations", where the quantities $\delta\eta^i(x)$ replace the finite-dimensional differentials of coordinates: $dx^i \to \delta\eta^i(x)$.

A *local functional* is a quantity which is written in local coordinates as

$$S\{f\} = \left(\int \cdots \int\right)_X L(x, f(x), f_j(x), \dots, f_{j_1 \dots j_m}(x))\, d^n x,$$
$$d^n x = dx^1 \wedge \cdots \wedge dx^n,$$

where $f = (f^1, \dots, f^k)$ is a local notation for a mapping $X \to Y$, and the $f^{(q)}_{j_1 \dots j_j}(x)$ are the partial derivatives of order $q$ of $f = (f^1, \dots, f^k)$ with respect to $(x^1, \dots, x^n)$ taken at some point $x \in X$, i.e., $f_j = \frac{\partial f}{\partial x^j}$, etc.

The integral is taken over the manifold $X$, possibly with a boundary $\partial X$. In the classical case, $X = \mathbb{R}^n$ or $X = D \subset \mathbb{R}^n$ is a domain with a smooth boundary.

We have already considered such functionals for $n = 1$. For piecewise smooth paths $y(t) = (y^1(t), \dots, y^k(t))$ in a domain in $\mathbb{R}^n$ with metric $dl^2 = g_{ij}\,dy^i\,dy^j$, we have introduced the length and action functionals $l(y) = \int_0^1 \sqrt{g_{ij}\dot{y}^i\dot{y}^j}\,dt$ and $S(y) = \int_0^1 g_{ij}\dot{y}^i\dot{y}^j\,dt$ (here $0 \le t = x^1 \le 1$). The critical points of these functionals are geodesic lines of the metric $g_{ij}\,dy^i\,dy^j$.

In this chapter we will consider multidimensional functionals, $n > 1$.

We will restrict ourselves to functionals $I$ whose Lagrangians depend only on $f$ and its first-order derivatives, but not on the derivatives of higher order. Such functionals already provide a large class of examples important for physics.

EXAMPLE. SURFACE AREA. Let $D \in \mathbb{R}^2$ be a domain in the plane with coordinates $x$ and $y$, let $f(x,y) = (u^1(x,y), u^2(x,y), u^3(x,y))$ be a two-dimensional surface in $\mathbb{R}^3$, and let $dl^2 = E\,dx^2 + 2F\,dx\,dy + G\,dy^2$ be the induced metric on the surface. The area functional of the two-dimensional surface is

$$I[f] = \iint_D \sqrt{EG - F^2}\,dx\,dy = \iint_D \sqrt{\langle f_x, f_x\rangle\langle f_y, f_y\rangle - \langle f_x, f_y\rangle^2}\,dx\,dy.$$

Now let $I$ be a generic functional with domain $F$, and let $f \in F$ be a point in $F$. Consider a smooth mapping $\eta$ from $D$ into $\mathbb{R}^k$ which is compactly supported, i.e., it vanishes outside an open set contained in $D$ together with its compact closure. In particular, we set $\eta = 0$ on the boundary $\partial D$ of $D$. The function $\eta$ is called a perturbation of $f$ if $f + \varepsilon\eta \in F$ for sufficiently small $\varepsilon$. Consider the derivative of the function $\varphi(\varepsilon) = I[f + \varepsilon\eta]$ at the point $\varepsilon = 0$:

$$\frac{dI[f+\varepsilon\eta]}{d\varepsilon}\bigg|_{\varepsilon=0} = \int_D \frac{\delta I}{\delta f}\,\delta f\,d^n x, \quad \delta f = \eta. \tag{14.1}$$

It is natural to call it "the derivative of the functional $I$ at the point $f$ in the direction $\eta$". The vector

$$\frac{\delta I}{\delta f} = \left(\frac{\delta I}{\delta f^1}, \dots, \frac{\delta I}{\delta f^k}\right),$$

defined by (14.1), is referred to as the *variational derivative of the functional $I[f]$*. A function $f_0 \in F$ is said to be a *stationary* (or *extremal, or critical*) function for the functional $I$ if

$$\frac{\delta I[f_0]}{\delta f} \equiv 0,$$

i.e., the expression (14.1) vanishes for any perturbation of the function $f$. Naturally, we can restrict our consideration to perturbations with sufficiently small supports.

**Theorem 14.1.** *A function $f_0 \in F$ is a stationary point of a functional $I[f]$ if and only if it satisfies the system of equations*

$$\frac{\delta I[f]}{\delta f^\alpha} = \frac{\partial L}{\partial f^\alpha} - \sum_{i=1}^{n} \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\Big) = 0 \quad (1 \le \alpha \le k). \tag{14.2}$$

Equations (14.2) are called the *Euler–Lagrange equations* for the functional $I$.

**Proof.** The proof consists in analytic calculation of the variational derivative and equating it to zero. We will show that the $\frac{\delta I[f]}{\delta f^\alpha}$ given by (14.2) constitute the vector $\frac{\delta I}{\delta f}$.

Consider the variation of the functional $I$:

$$\delta I[f] = \int_D [L(x^i, f^\alpha + \varepsilon\eta^\alpha, f^\beta_{x^j} + \varepsilon\eta^\beta_{x^j}) - L(x^i, f^\alpha, f^\beta_{x^j})]\, d^n x.$$

Write down the Taylor expansion of the integrand:

$$\begin{aligned}\delta I[f] &= \int_D \Big[\sum_{\alpha=1}^{k} \frac{\partial L}{\partial f^\alpha}\,\varepsilon\eta^\alpha + \sum_{\alpha=1}^{k}\sum_{i=1}^{n} \frac{\partial L}{\partial f^\alpha_{x^i}}\,\varepsilon\eta^\alpha_{x^i} + o(\varepsilon)\Big]\, d^n x\\ &= \varepsilon \int_D \sum_{\alpha=1}^{k} \Big[\frac{\partial L}{\partial f^\alpha}\,\eta^\alpha + \sum_{i=1}^{n} \frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha_{x^i}\Big]\, d^n x + \int_D o(\varepsilon)\, d^n x.\end{aligned}$$

Integrating by parts, we obtain

$$\begin{aligned}\delta I[f] &= \varepsilon \int_D \sum_{i=1}^{n} \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha\Big)\, d^n x\\ &\quad + \varepsilon \int_D \sum_{\alpha=1}^{k}\Big(\frac{\partial L}{\partial f^\alpha} - \sum_{i=1}^{n} \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\Big)\Big)\eta^\alpha\, d^n x + \int_D o(\varepsilon)\, d^n x.\end{aligned}$$

In the first term the integration over the variable $x^i$ can be separated from the integration over the other variables $x^1, \ldots, \hat{x}^i, \ldots, x^n$. We obtain

$$\int_D \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha\Big)\, d^n x = \int_{x^1, \ldots, \widehat{x^i}, \ldots, x^n} \Big[\int_P^Q \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha\Big)\, dx^i\Big]\, d^{n-1}x,$$

where $P$ and $Q$ lie on the boundary of $D$ and depend on $x^1, \ldots, \widehat{x^i}, \ldots, x^n$, while $d^{n-1}x = dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n$. In the inner integral $\int_P^Q$ the

variables $x^1, \dots, \widehat{x^i}, \dots, x^n$ may be regarded as parameters; by integration over $x^i$ we obtain

$$\int_D \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha\Big)\,d^n x = \int_{x^1,\dots,\widehat{x^i},\dots,x^n}\Big[\frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha\Big|_Q - \frac{\partial L}{\partial f^\alpha_{x^i}}\,\eta^\alpha\Big|_P\Big]\,d^{n-1}x \equiv 0,$$

since $\eta^\alpha(Q) = \eta^\alpha(P) = 0$ by the choice of $\eta$. Thus we have

$$\delta I[f] = \varepsilon \int_D \sum_{\alpha=1}^{k}\Big[\frac{\partial L}{\partial f^\alpha} - \sum_{i=1}^{n}\frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\Big)\Big]\eta^\alpha\,d^n x + \int_D o(\varepsilon)\,d^n x.$$

Since $\int_D o(\varepsilon)\,d^n x = o(\varepsilon)$, we obtain

$$\frac{\delta I[f]}{\delta f^\alpha} = \frac{\partial L}{\partial f^\alpha} - \sum_{i=1}^{n}\frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\Big). \tag{14.3}$$

Hence the theorem. □

Formula (14.3) gives an explicit expression for the *variational derivative* of the functional $I$.

Note that the definition of a stationary point of a functional is of general nature, and calculations similar to those performed above, but more cumbersome, yield the Euler–Lagrange equations for functionals with Lagrangians depending on higher order derivatives.

**14.1.2. Energy-momentum tensor and conservation laws.** Conservation laws of a physical system arise from symmetries of this system. For example, if the Lagrangian $L$ of a univariate variational problem does not depend on time explicitly, then the energy of the system is conserved. In case such Lagrangian is independent of some spacial variable $x^i$, the corresponding momentum is conserved:

$$\frac{dp_i}{dt} = \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}^i}\Big) = 0.$$

This correspondence between symmetries and conservation laws is explained by the following theorem.

**Theorem 14.2** (Noether). *Let*

$$x \to \tilde{x}(x,\tau), \quad f^\alpha \to \tilde{f}^\alpha(x, f^\alpha, \tau) \tag{14.4}$$

*be a transformation of coordinates $x^i$ and field variables $f^\alpha$ specified by smooth functions of finitely many parameters $\tau_1, \dots, \tau_s$. Assume that the variation of the action $\int_D L\,d^n x$ relative to this transformation equals zero for any domain $D$. Then there exist $s$ conservation laws (dynamical invariants), i.e., there exist $s$ functionals of field variables which do not change with time and are in one-to-one correspondence with deformations related to the parameters $\tau_1, \dots, \tau_s$.*

The proof of this theorem consists in explicit construction of these invariants based on deformations (14.4). To each parameter $\tau_\mu$ corresponds the *current vector* $J^k_\mu$ with zero divergence:

$$\frac{\partial J^k_\mu}{\partial x^k} = 0. \tag{14.5}$$

Consider the case where the fields $f^\alpha$ are defined in the Minkowski space $\mathbb{R}^{1,3}$ with coordinates $x^0 = ct, x^1, x^2, x^3$. Then (14.5) implies that the quantity

$$F_\mu(t) = \int_{t=\text{const}} J^k_\mu \, dS_k$$

is preserved. Here $dS_k = \frac{1}{6}\,\varepsilon_{ijlk}\, dx^i \wedge dx^j \wedge dx^l$.

We will present this scheme in an important example where the Lagrangian does not depend on the spacial variables $x^i$, $0 \le i \le 3$. Then the functional has the form

$$I[f] = \int L(f^\alpha, f^\beta_{x^k})\, d^4x.$$

The continuous symmetries form the Poincaré group,

$$x \to Ax + b, \quad A \in \mathrm{O}(1,3),\ b \in \mathbb{R}^{1,3},$$

which preserves the Minkowski metric

$$ds^2 = g_{ik}\, dx^i\, dx^k = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2.$$

The extremal functions $f$ describing the "motion of the system" satisfy the system of Euler–Lagrange equations

$$\frac{\partial L}{\partial f^\alpha} - \sum_i \frac{\partial}{\partial x^i}\Big(\frac{\partial L}{\partial f^\alpha_{x^i}}\Big) = 0.$$

The functional $I$ is invariant under transitions to other inertial frames: $x \to x' = Ax + b$. Under such a transition a scalar $f^\alpha$ goes into $\tilde{f}^\alpha(x') = f^\alpha(x)$, and if the $f^\alpha$ form a vector $f^i$, then the symmetry transformations have the form

$$x^i \to x' = a^i_j x^j + b^i, \quad f^i(x) \to \tilde{f}^i(x') = a^i_j f^j(x), \quad A = (a^i_j) \in \mathrm{O}(1,3).$$

In general, tensor fields are transformed by the tensor transformation rules.

The "laws of motion" specified by the Euler–Lagrange equations for the functional $I$ are the same at all points of the "space". This fact is basic for the derivation of conservation laws, to which we now proceed.

We derive the conservation laws corresponding to the translations

$$x \to x' = x + \xi, \quad f^\alpha(x) \to \tilde{f}^\alpha(x') = f^\alpha(x), \quad d^4x' = d^4x.$$

Suppose we are given a solution $f^\alpha(x)$ to the Euler–Lagrange equations in $\mathbb{R}^{1,3}$. For an arbitrary domain $U \subset \mathbb{R}^{1,3}$, consider variations of the action generated by translations

$$x \to x' = x + s\xi, \quad \xi \in \mathbb{R}^{1,3}, \quad s \in \mathbb{R}.$$

They are

$$\delta I = \int_{U+s\xi} L(\tilde{f}^\alpha(x'), \tilde{f}^\beta_{x'^i}(x'))\, d^4x' - \int_U L(f^\alpha(x), f^\beta_{x^i}(x))\, d^4x \equiv 0,$$

where $U + s\xi$ is the domain $U$ shifted by $s\xi$. Now we write the equality $\delta I = 0$ explicitly, with this variation evaluated for the given solution $f^\alpha$ to the Euler–Lagrange equations:

$$\delta I = s\int_U \frac{\partial L}{\partial x^i}\,\xi^i\, d^4x + s\int_U \Big(\frac{\partial L}{\partial f^\alpha}\,\delta_i f^\alpha + \frac{\partial L}{\partial f^\alpha_{x^k}}\,\delta_i f^\alpha_{x^k}\Big)\xi^i\, d^4x + o(s) \equiv 0. \tag{14.6}$$

The first term here is due to the change of the domain $U$, and by $\frac{\partial L}{\partial x^i}$ we mean the derivative of the Lagrangian taken on the given solution: $L = L(f^\alpha(x), f^\beta_{x^k}(x))$.

The second term is related to the change of the form of the functions $f^\alpha$. Indeed, we have a family of functions $\tilde{f}^\alpha(\tilde{x}) = g^\alpha(s, x + s\xi) \equiv f^\alpha(x)$ parametrized by $s$, and by $\xi^i\delta_i f^\alpha$ we mean the derivative with respect to the first argument $s$ for $s = 0$. In particular, for a basis vector $\xi = e_i$, where $\xi^j = \delta^j_i$, we have

$$\frac{\partial \tilde{f}^\alpha}{\partial s}\Big|_{s=0} = \delta_i f^\alpha + \frac{\partial f^\alpha}{\partial x^i} \equiv 0.$$

Substituting the expressions

$$\delta_i f^\alpha = -\frac{\partial f^\alpha}{\partial x^i}, \quad \delta_i f^\alpha_{x^k} = -\frac{\partial f^\alpha_{x^k}}{\partial x^i}$$

into equation (14.6), we obtain

$$\delta I = s\int_U \Big(\frac{\partial L}{\partial x^i} - \frac{\partial L}{\partial f^\alpha}\frac{\partial f^\alpha}{\partial x^i} - \frac{\partial L}{\partial f^\alpha_{x^k}}\frac{\partial f^\alpha_{x^k}}{\partial x^i}\Big)\xi^i\, d^4x + o(s) \equiv 0.$$

Now we successively substitute the Euler–Lagrange equations (14.2) and the equality $\frac{\partial}{\partial x^i}(f^\alpha_{x^k}) = \frac{\partial^2 f^\alpha}{\partial x^i \partial x^k} = \frac{\partial}{\partial x^k}(f^\alpha_{x^i})$ into the integrand:

$$\begin{aligned}
\delta I &= s\int_U \Big(\frac{\partial L}{\partial x^i} - \frac{\partial}{\partial x^k}\Big(\frac{\partial L}{\partial f^\alpha_{x^k}}\Big)\frac{\partial f^\alpha}{\partial x^i} - \frac{\partial L}{\partial f^\alpha_{x^k}}\frac{\partial (f^\alpha_{x^k})}{\partial x^i}\Big)\xi^i\, d^4x + o(s) \\
&= s\int_U \Big(\frac{\partial L}{\partial x^i} - \frac{\partial}{\partial x^k}\Big(\frac{\partial L}{\partial f^\alpha_{x^k}}\Big)\frac{\partial f^\alpha}{\partial x^i} - \frac{\partial L}{\partial f^\alpha_{x^k}}\frac{\partial}{\partial x^k}(f^\alpha_{x^i})\Big)\xi^i\, d^4x + o(s) \\
&= s\int_U \Big(\frac{\partial L}{\partial x^i} - \frac{\partial}{\partial x^k}\Big(\frac{\partial L}{\partial f^\alpha_{x^k}}\, f^\alpha_{x^i}\Big)\Big)\xi^i\, d^4x + o(s) \equiv 0.
\end{aligned}$$

Since the domain $U$ and the vector $\xi$ are arbitrary, we conclude that for any solution to the Euler–Lagrange equations the following equality holds:

$$\frac{\partial L}{\partial x^i} = \delta_i^k \frac{\partial L}{\partial x^k} = \frac{\partial}{\partial x^k}\Big(\frac{\partial L}{\partial f^\alpha_{x^k}} f^\alpha_{x^i}\Big).$$

We rewrite it as

$$\frac{\partial}{\partial x^k}\Big(\delta_i^k L - f^\alpha_{x^i}\frac{\partial L}{\partial f^\alpha_{x^k}}\Big) = 0. \tag{14.7}$$

The tensor

$$T_i^k = f^\alpha_{x^i}\frac{\partial L}{\partial f^\alpha_{x^k}} - \delta_i^k L \tag{14.8}$$

is called the *energy-momentum tensor* of a system with Lagrangian $L(f, f_{x^k})$. It is written also as follows:

$$T_{ik} = g_{kl}T_i^l, \quad T^{ik} = g^{kl}T_l^i.$$

Relation (14.7) means that the divergence of the tensor $T_i^k$ is everywhere equal to zero:

$$\frac{\partial T_i^k}{\partial x^k} = 0. \tag{14.9}$$

Note that the above derivation is valid for any translation-invariant Lagrangian $L$ in $\mathbb{R}^n$.

In the special case where $n = 1$, $x^1 = t$, and $f^\alpha = y^\alpha$ are local coordinates in some manifold $M^k$, we obtain a one-dimensional variational problem. In this case the energy-momentum tensor consists of one component

$$T = T_1^1 = \dot y^\alpha \frac{\partial L}{\partial \dot y^\alpha} - L = E,$$

which is the energy of the system. Relation (14.9) is simply the energy conservation law (see 12.3.1):

$$\frac{dE}{dt} = 0.$$

Let us introduce the 3-forms

$$\omega_i = \sum_k (-1)^{k-1} T_i^k\, dx^0 \wedge \cdots \wedge \widehat{dx^k} \wedge \cdots \wedge dx^3 = T_i^k\, dS_k.$$

Since

$$d\omega_i = \Big(\sum_k \frac{\partial T_i^k}{\partial x^k}\Big)\, dx^0 \wedge \cdots \wedge dx^3,$$

by the Stokes theorem relations (14.9) are equivalent to the property that

$$\int_{\partial D} \omega_i = 0, \quad i = 0, \dots, 3, \tag{14.10}$$

for any domain $D$ with piecewise smooth boundary.

There is a simpler way to derive relations (14.9): for a given solution $(f^\alpha)$ it suffices to write down the derivatives of the Lagrangian $L$ with respect to $x^i$,

$$\frac{\partial L}{\partial x^i} - \frac{\partial L}{\partial f^\alpha}\frac{\partial f^\alpha}{\partial x^i} - \frac{\partial L}{\partial f^\alpha_{x^k}}\frac{\partial (f^\alpha_{x^k})}{\partial x^i} = 0$$

and, as was done before, substitute the Euler–Lagrange equations and the equality $\frac{\partial}{\partial x^i}(f^\alpha_{x^k}) = \frac{\partial^2 f^\alpha}{\partial x^i \partial x^k} = \frac{\partial}{\partial x^k}(f^\alpha_{x^i})$ into the left-hand side to obtain (14.9). We derived these formulas in the framework of a more general scheme, which is applicable also to derivation of conservation laws related to rotation of the space.

Now we will show why relations (14.9) themselves are sometimes called "conservation laws".

Consider the vector $P = (P_0, P_1, P_2, P_3)$ (with subscripts) of the form

$$P_i = \frac{1}{c}\int_{x^0=\text{const}} T_i^k \, dS_k, \quad i = 0, 1, 2, 3,$$

where $c$ is the velocity of light in vacuum. It is called the *momentum* 4-*vector* (with subscripts) of the system with Lagrangian $L$.

By analogy with formula (12.10) for the energy, the density of the energy is defined as the component

$$T_{00} = \dot{f}^\alpha \frac{\partial L}{\partial \dot{f}^\alpha} - L,$$

where the dot denotes differentiation with respect to $x^0 = ct$. This means, in particular, that the energy is localized and the total energy of the system equals

$$\int T_{00} \, dx^1 \wedge dx^2 \wedge dx^3.$$

Recall that the energy of a relativistic particle is also equal to $cP_0$, where $P$ is the momentum 4-vector of the particle (Section 12.3).

Naturally, we assume everywhere that the fields $f$ decay at infinity sufficiently fast, all the integrals involved are convergent, and the momentum 4-vector of the system is well defined.

**Theorem 14.3.** *The identity*

$$\frac{\partial T_i^k}{\partial x^k} = 0$$

*implies that the momentum* 4-*vector* $P$ *is conserved in time* (*does not depend on* $x^0 = ct$).

**Proof.** Consider the cylinder $C$ in the Minkowski space with circular bases $D_1$ and $D_2$ of sufficiently large radius $R$ lying in hyperplanes $x^0 = x^0_1$ and $x^0 = x^0_2$. Denote by $\Pi$ the lateral surface of the cylinder.

Equalities (14.10) imply that $\int_C T_i^k\,dS_k = 0$ for each $i$. Hence

$$\Big(\int_{D_2} - \int_{D_1} + \int_{\Pi}\Big) T_i^k\,dS_k = 0.$$

As the radius $R$ of the cylinder tends to infinity, the integral over its lateral surface $\Pi$ tends to zero, and we obtain

$$P_i(x_1^0) = \int_{x^0 = x_1^0} T_i^k\,dS_k = \int_{x^0=x_2^0} T_i^k\,dS_k = P_i(x_2^0).$$

Hence the theorem. □

Each tensor $\psi_i^{kl}$ that is skew-symmetric in superscripts $k$ and $l$ specifies the transformation of the tensor $T_i^k$ by the formula

(14.11) $$T_i^k \to T_i^k + \frac{\partial \psi_i^{kl}}{\partial x^l}, \quad \text{where } \psi_i^{kl} = -\psi_i^{lk}.$$

The new tensor also satisfies the equation $\frac{\partial T_i^k}{\partial x^k} = 0$, since $\frac{\partial^2 \psi_i^{kl}}{\partial x^k \partial x^l} = 0$. Assume that the tensor $\psi_i^{kl}$ decays sufficiently fast at infinity along with its derivatives.

**Lemma 14.1.** *The momentum 4-vector of the system remains unchanged under the transformation* (14.11).

**Proof.** We must show that $\int_{x^0=\text{const}} \frac{\partial \psi_i^{kl}}{\partial x^l}\,dS_k = 0$. Since the tensor $\psi_i^{kl}$ is skew-symmetric in $k, l$, we have

$$\int_{x^0=\text{const}} \frac{\partial \psi_i^{kl}}{\partial x^l}\,dS_k = \frac{1}{2}\int_{x^0=\text{const}} \Big(dS_k\,\frac{\partial \psi_i^{kl}}{\partial x^l} - dS_l\,\frac{\partial \psi_i^{kl}}{\partial x^k}\Big) = \int_{x^0=\text{const}} d\eta_i,$$

where $\eta_i$ is a 2-form on the surface $x^0 = \text{const}$. By the Stokes theorem, the integral of $d\eta_i$ over the surface $x^0 = \text{const}$ is equal to the integral of $\eta_i$ over a two-dimensional sphere "inflated to infinity" and "enclosing" the surface $x^0 = 0$. Since the form $\eta_i$ decreases at infinity sufficiently fast, this integral is equal to zero. Hence the lemma. □

The energy-momentum tensor as such is of no physical interest, and only invariants of type $P_i$ defined by means of this tensor are important. For this reason it is defined up to transformations that preserve these invariants.

In general, the tensor $T^{ik}$ is not symmetric, but it can be made so with the aid of transformation (14.11). We will call the tensor thus obtained again the energy-momentum tensor, since it satisfies equation (14.9) and gives rise to the same momentum 4-vector as the tensor (14.8). In examples of physical importance one can make the symmetrized tensor $T^{ik}$ decay at infinity also sufficiently fast.

The rotations of the space give rise to the dynamical invariant, which is called the *moment of momentum*. It is obtained according to the general scheme, by integration of the tensor $S^l_{ik}$ satisfying the "conservation law" $\frac{\partial S^l_{ik}}{\partial x^l} = 0$ over the surfaces $x^0 = \text{const}$.

Suppose the tensor $T^{ik}$ is symmetric. Then the moment of momentum has a particularly simple form: this is the tensor

$$M^{ik} = \int (x^i\,dP^k - x^k\,dP^i) = \frac{1}{c}\int_{x^0=\text{const}} (x^i T^{kl} - x^k T^{il})\,dS_l$$

(for convenience we define this tensor with superscripts raised by means of the Minkowski metric). This formula generalizes the one for the moment of momentum of a system of particles in classical mechanics (see Section 12.3):

$$M^{ik} = \sum (x^i P^k - x^k P^i),$$

where the summation extends over all particles of the system.

Symmetrization of the tensor $T^{ik}$ is not defined uniquely, but up to addition of the term

$$T^{ik} \to T^{ik} + \frac{\partial \varphi^{ikl}}{\partial x^l}, \quad \text{where } \varphi^{ikl} = \varphi^{kil} = -\varphi^{ilk}.$$

As before, we require the tensor $\varphi^{ikl}$ to decay sufficiently fast at infinity along with its derivatives. Then it can be shown in a similar way to the proof of Lemma 14.1 that the addition of $\frac{\partial \varphi^{ikl}}{\partial x^l}$ does not affect the moment of momentum.

**Theorem 14.4.** *The moment-of-momentum tensor $M^{ik}$ is preserved in time.*

**Proof.** We have

$$M^{ik}(x^0) = \frac{1}{c}\int_{x^0=\text{const}} (x^i T^{kl} - x^k T^{il})\,dS_l.$$

Similarly to the proof of Theorem 14.3 it suffices to show that

$$\frac{\partial}{\partial x^l}(x^i T^{kl} - x^k T^{il}) = 0.$$

Taking the derivatives, we obtain

$$\frac{\partial}{\partial x^l}(x^i T^{kl} - x^k T^{il}) = T^{ki} - T^{ik} = 0$$

by symmetry of the tensor $T^{ik}$ and due to the equation $\frac{\partial T^{kl}}{\partial x^l} = 0$, which follows from (14.9) and the fact that the coefficients of the metric $g_{ij}$ are constant. Hence the theorem. □

We considered the definition of the energy-momentum tensor in the case where the metric $g_{ij}$ has constant coefficients. In the general case of a curved space the energy-momentum tensor is defined in a different way and, in particular, is symmetric by definition. We will present it in Section 15.1.

## 14.2. Examples of multidimensional variational problems

**14.2.1. Minimal surfaces.** In Section 4.5 we considered minimal surfaces in the three-dimensional Euclidean space. They were defined as surfaces of zero mean curvature. At the same time the very term indicates that they are related to variational problems. Here we will show that they are critical points of the area functional, which we have already treated as an example in 14.1.1.

Let $\Sigma$ be an oriented surface in $\mathbb{R}^3$ specified locally by a regular mapping

$$r : U \to \mathbb{R}^3$$

of a domain $U$ with coordinates $x^1, x^2$ into the three-dimensional space. On this surface the surface area form,

$$d\sigma = \sqrt{\det(g_{ij})}\, dx^1 \wedge dx^2 = \sqrt{g_{11}g_{22} - g_{12}^2}\, dx^1 \wedge dx^2,$$

is defined, where in each tangent space the frame $r_1, r_2$ corresponding to the coordinates $x^1, x^2$ is positively oriented and

$$g_{ij}\, dx^i\, dx^j = E(dx^1)^2 + 2F\, dx^1\, dx^2 + G(dx^2)^2$$

is the first fundamental form of the surface.

A one-parameter family of surfaces $\Sigma_\varepsilon$ is called a smooth deformation of $\Sigma$ if:

1) $\Sigma_0 = \Sigma$;

2) the surfaces $\Sigma_\varepsilon$ are specified by mappings $r^\varepsilon$, which are smooth functions of the deformation parameter $\varepsilon$.

A closed set $V \subset \Sigma$ is called a deformation support if the part of the surface lying outside $V$ is not affected by the deformation.

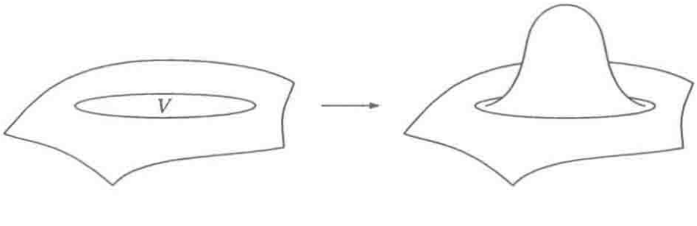


**Figure 14.1.** Deformation support.

If the deformation support $V$ is compact, then the area $S(\varepsilon)$ of the deformed part $V_\varepsilon$ is finite and is a smooth function of $\varepsilon$. A surface is said

to be *minimal* if it is a critical point of the area functional relative to all deformations with compact support:

$$\frac{d}{d\varepsilon} S(\varepsilon)\Big|_{\varepsilon=0} = 0.$$

In particular, if a surface that spans a contour in $\mathbb{R}^3$ has the minimum area among all surfaces encircled by this contour, then this surface is minimal. But in general the minimal surface does not minimize the area functional. This is in analogy with the case of geodesics, which are "one-dimensional minimal surfaces".

**Theorem 14.5.** *A regular surface $\Sigma$ specified by a mapping $r\colon U \to \mathbb{R}^3$ is minimal if and only if its mean curvature is everywhere equal to zero,*

$$H = 0.$$

**Proof.** Let $V$ be a closed subset of $U$ and let $r^\varepsilon : U \to \mathbb{R}^3$ be a deformation of the surface with support in $V$. It may be expanded into a series in $\varepsilon$, and in this expansion we may neglect translation deformations along the surface (because they preserve the area). Then we obtain

$$r^\varepsilon(x^1, x^2) = r(x^1, x^2) + \varepsilon\varphi\mathbf{n} + O(\varepsilon^2),$$

where the function $\varphi$ vanishes outside $V$ and $\mathbf{n}$ is the unit normal to the surface. The area of the deformed part of the surface $r^\varepsilon(V)$ equals

$$S(\varepsilon) = \int_V \sqrt{(r_1^\varepsilon, r_1^\varepsilon)(r_2^\varepsilon, r_2^\varepsilon) - (r_1^\varepsilon, r_2^\varepsilon)(r_1^\varepsilon, r_2^\varepsilon)}\, dx^1 \wedge dx^2.$$

Since

$$r_k^\varepsilon = r_k + \varepsilon\varphi\mathbf{n}_k + \varepsilon\varphi_k\mathbf{n} + O(\varepsilon^2)$$

and $(r_1, \mathbf{n}) = (r_2, \mathbf{n}) = 0$, we obtain

$$(r_i^\varepsilon, r_j^\varepsilon) = (r_i, r_j) + \varepsilon\varphi((r_i, \mathbf{n}_j) + (r_j, \mathbf{n}_i)) + O(\varepsilon^2)$$

(the subscripts $i$ mean differentiation with respect to $x^i$). The derivational equations (see 3.4.1) imply that $(r_i, \mathbf{n}_j) = -b_{ij}$, where $b_{ij}\, dx^i\, dx^j$ is the second fundamental form of the surface. Hence we see that

$$\begin{aligned} S(\varepsilon) &= \int_V \sqrt{1 - 2\varepsilon\varphi\frac{b_{11}g_{22} + b_{22}g_{11} - b_{12}g_{21} - b_{21}g_{12}}{g_{11}g_{22} - g_{12}^2} + O(\varepsilon^2)}\, d\sigma \\ &= S(0) - \varepsilon\int_V \frac{b_{11}g_{22} + b_{22}g_{11} - b_{12}g_{21} - b_{21}g_{12}}{g_{11}g_{22} - g_{12}^2}\varphi\, d\sigma + O(\varepsilon^2). \end{aligned}$$

One can easily check that the sum of the roots $k_1$ and $k_2$ of the equation $P(\lambda) = \det(b_{ij} - \lambda g_{ij}) = 0$ is equal to

$$k_1 + k_2 = \frac{b_{11}g_{22} + b_{22}g_{11} - b_{12}g_{21} - b_{21}g_{12}}{g_{11}g_{22} - g_{12}^2},$$

and by Theorem 3.4, this is twice the mean curvature of the surface, $2H = k_1 + k_2$.

Thus we obtain the following formula for the variation of the area functional:

$$\frac{d}{d\varepsilon}S(\varepsilon)\Big|_{\varepsilon=0} = -2\int_V H\varphi\, d\sigma. \tag{14.12}$$

A surface is a critical point of the area functional, i.e., the variation of the area vanishes for any deformation $V$ with compact support, if the right-hand side of (14.12) equals zero for any compactly supported smooth function $\varphi$. This holds if and only if $H = 0$ at each point of the surface. Hence the theorem. □

**14.2.2. Electromagnetic field equations.** We stated the Maxwell equations without deriving them in 9.1.3. Here we will derive these equations from the variational principles.

The action of a system consisting of the electromagnetic field and charged particles in this field has the form

$$S = S_m + S_{fm} + S_f.$$

The term $S_m$ is determined by moving charged particles, which we regard as point particles, and it has the same form as the action for the system of particles in the absence of the field:

$$S_m = -\sum mc\int_a^b dl.$$

Here the summation is extended over all particles, $m$ is the charge of a particle, $dl$ is the length element in the Minkowski metric in $\mathbb{R}^{1,3}$, and the integral is taken along the world line of the particle between two fixed events. The three-dimensional expression for this term is

$$S_m = -\int_{t_1}^{t_2} mc^2\sqrt{1-\frac{v^2}{c^2}}\, dt,$$

where $v$ is the three-dimensional velocity of the particle and the integration is along its trajectory between the initial and final time values (see Section 12.3).

The term $S_{mf}$ is due to the interaction between particles and the field and has the form

$$S_{mf} = -\sum \frac{e}{c}\int A_k\, dx^k,$$

where the summation is extended over all particles, $e$ is the charge of a particle, and the integral is taken along the world line of the particle. Here $A_i$ is the 4-potential of the field, which determines the properties of the field. The expression $A_i\, dx^i$ specifies a 1-form in the Minkowski space $\mathbb{R}^{1,3}$, and

its coefficients depend on time and spacial coordinates. The properties of a particle from the viewpoint of its interaction with the electromagnetic field are determined solely by its charge $e$.

The term $S_f$ depends on the properties of the field itself; it is the action of the field in the absence of charges.

In order to obtain the equations describing this system, we must write down the Euler–Lagrange equations for the complete functional $S$ with varying trajectories of the particles (for fixed boundary conditions) and varying field. Since the term $S_f$ does not change under variation of trajectories, in order to derive the equations of motion of charged particles in the given electromagnetic field, we need to write down the Euler–Lagrange equations for the functional

$$S_m + S_{fm} = \sum \int_a^b \Big(-mc\,dl - \frac{e}{c}\,A_k\,dx^k\Big).$$

The spacial components $(A^1, A^2, A^3)$ of the 4-vector $A$ form the *vector potential* $\mathbf{A}$ of the field, and the time component $A^0 = \varphi$ is the *scalar potential* of the field. The superscripts are raised here by means of the Minkowski metric: $A = (\varphi, -\mathbf{A})$.

The exterior derivative of the form $A_i\,dx^i$ is called the *electromagnetic field tensor* $F_{ik}$:

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -H_z & H_y \\ -E_y & H_z & 0 & -H_x \\ -E_z & -H_y & H_x & 0 \end{pmatrix},$$

where $x^0 = ct,\ x^1 = x,\ x^2 = y,\ x^3 = z$. The vector

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \operatorname{grad}\varphi$$

is called the *electric field strength*, and the vector

$$\mathbf{H} = \operatorname{curl}\mathbf{A}$$

is the *magnetic field strength.* An electromagnetic field with $\mathbf{E} \neq 0$, $\mathbf{H} = 0$ is called an *electric field*, and a field with $\mathbf{E} = 0$, $\mathbf{H} \neq 0$ is a *magnetic field.* The obvious equation

$$d(F_{ik}\,dx^i \wedge dx^k) = d^2(A_i\,dx^i) = 0$$

gives the first pair of the *Maxwell equations*:

$$\begin{aligned} \operatorname{curl}\mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}, \\ \operatorname{div}\mathbf{H} &= 0. \end{aligned} \tag{14.13}$$

We set the action $S_f$ to be

$$S_f = -\frac{1}{16\pi c}\int 2(H^2 - E^2)\,d^4x,$$

where $H^2 = \langle \mathbf{H}, \mathbf{H}\rangle$, $E^2 = \langle \mathbf{E}, \mathbf{E}\rangle$ are scalar squares and integration is extended over the entire three-dimensional space for the spacial coordinates and between two fixed time values for $x^0 = ct$. Since $F_{ik}^2 \equiv F_{ik}F^{ik} = 2(H^2 - E^2)$, the action $S_f$ becomes

$$S_f = -\frac{1}{16\pi c}\int F_{ik}^2\,d^4x.$$

The total action $S$ for the electromagnetic field with charges inside can be written as

$$S = -\sum\int mc\,dl - \sum\int \frac{e}{c}A_k\,dx^k - \frac{1}{16\pi c}\int F_{ik}^2\,d^4x.$$

In order to obtain the electromagnetic field equations, we must write down the Euler–Lagrange equations for the functional $S$ with only the potentials of the field varying.

We represent the functional $S_f + S_{fm}$ as the integral over the entire space. To this end we introduce the charge density $\rho$, which depends on time and coordinates $x^1, x^2, x^3$. Then $\rho\,dV$ is the charge in the three-dimensional volume $dV$. If $\frac{dx^i}{dt}$ is the velocity 4-vector of the point charge, then the current 4-vector is $j = (c\rho, \mathbf{j})$, where $j^i = \rho\frac{dx^i}{dt}$, $\mathbf{j} = (j^1, j^2, j^3) = \rho\mathbf{v}$, and $\mathbf{v}$ is the velocity of the charge at a given point. The vector $\mathbf{j}$ is called the current 3-vector. The action becomes

$$S = -\sum\int mc\,dl - \frac{1}{c^2}\int A_i j^i\,d^4x - \frac{1}{16\pi c}\int F_{ik}^2\,d^4x.$$

Now we proceed to derivation of the field equations. Since the trajectories of the charges are not subject to variation, we obtain that $\delta S_m = 0$ and in the term $S_{mf}$ there is no variation of the current $j^i$. We have

$$\begin{aligned}\delta S = \delta(S_{mf} + S_f) &= -\int\Big(\frac{1}{c^2}\,j^i\delta A_i + \frac{1}{16\pi c}\,\delta(F_{ik}^2)\Big)\,d^4x\\ &= -\int\frac{1}{c}\Big\{\frac{1}{c}\,j^i\delta A_i + \frac{1}{8\pi}\,F^{ik}\delta F_{ik}\Big\}\,d^4x = 0.\end{aligned}$$

Since $F_{ik} = -F_{ki} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}$, we obtain

$$\begin{aligned}\delta S &= -\int\frac{1}{c}\Big\{\frac{1}{c}j^i\delta A_i + \frac{1}{8\pi}\,F^{ik}\delta\Big(\frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}\Big)\Big\}\,d^4x\\ &= -\int\frac{1}{c}\Big\{\frac{1}{c}\,j^i\delta A_i + \frac{1}{8\pi}\,F^{ik}\frac{\partial}{\partial x^i}\,(\delta A_k) - \frac{1}{8\pi}\,F^{ik}\frac{\partial}{\partial x^k}\,(\delta A_i)\Big\}\,d^4x\\ &= -\int\frac{1}{c}\Big\{\frac{1}{c}\,j^i\delta A_i - \frac{1}{4\pi}\,F^{ik}\frac{\partial}{\partial x^k}\,(\delta A_i)\Big\}\,d^4x.\end{aligned}$$

The second integral is transformed using integration by parts:

$$\delta S = -\int_D \frac{1}{c}\Big\{\frac{1}{c}\, j^i \delta A_i + \frac{1}{4\pi}\,\delta A_i\,\frac{\partial F^{ik}}{\partial x^k}\Big\}\, d^4x - \frac{1}{4\pi c}\int_{\partial D} F^{ik}\,\delta A_i\, dS_k,$$

where, as we have pointed out before, the domain $D$ is specified by the inequalities $t_1 \le t \le t_2$. The fields are fast decreasing at infinity, and for $t = t_1, t_2$ the variations of the potential vanish, $\delta A_i = 0$. Hence the second term vanishes and we have

$$\delta S = \int \Big(\frac{1}{c} j^i + \frac{1}{4\pi}\,\frac{\partial F^{ik}}{\partial x^k}\Big)\delta A_i\, d^4x = 0.$$

Since the variations $\delta A_i$ inside the domain are arbitrary, we obtain

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c}\, j^i.$$

We rewrite these equations in the three-dimensional form. For $i = 1$ the equation is

$$\frac{1}{c}\,\frac{\partial F^{10}}{\partial t} + \frac{\partial F^{11}}{\partial x} + \frac{\partial F^{12}}{\partial y} + \frac{\partial F^{13}}{\partial z} = -\frac{4\pi}{c} j^1.$$

In terms of the vectors $\mathbf{E}$ and $\mathbf{H}$, we obtain $-\frac{\partial H_z}{\partial y} + \frac{\partial H_y}{\partial z} + \frac{1}{c}\frac{\partial E_x}{\partial t} = -\frac{4\pi}{c}\, j_x$. In a similar way we transform the equations for $i = 2, 3$ and rewrite these three equations as

$$\operatorname{curl}\mathbf{H} = \frac{1}{c}\,\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\,\mathbf{j}. \tag{14.14}$$

For $i = 0$ we have

$$\frac{\partial(E_x)}{\partial x} + \frac{\partial(E_y)}{\partial y} + \frac{\partial(E_z)}{\partial z} = \frac{4\pi}{c}\, c\rho,$$

i.e.,

$$\operatorname{div}\mathbf{E} = 4\pi\rho. \tag{14.15}$$

Equations (14.14) and (14.15) are the second pair of the Maxwell equations.

Note that the Maxwell equations (14.13), (14.14), and (14.15) have nontrivial solutions in the absence of charges as well (for $\rho = 0$, $\mathbf{j} = 0$). Such solutions describe the *electromagnetic waves*.

Let us find the energy-momentum tensor of the electromagnetic field if there is no charge. The Lagrangian of the action $S_f$ equals

$$L = -\frac{1}{16\pi c}\, F_{ik}^2 = -\frac{1}{16\pi c}\Big(\frac{\partial A_l}{\partial x^k} - \frac{\partial A_k}{\partial x^l}\Big)^2.$$

The field variables are the components of the potential $A_i$. Therefore, by the definition of the energy-momentum tensor we obtain

$$T_i^k = \frac{\partial A_l}{\partial x^i} \frac{\partial L}{\partial\left(\frac{\partial A_l}{\partial x^k}\right)} - \delta_i^k L.$$

Since the tensor $F^{kl}$ is skew-symmetric, we have

$$dL = -\frac{1}{8\pi c} F^{kl}\left(d\frac{\partial A_l}{\partial x^k} - d\frac{\partial A_k}{\partial x^l}\right) = -\frac{1}{4\pi c} F^{kl} d\frac{\partial A_l}{\partial x^k}.$$

Hence we conclude that $\frac{\partial L}{\partial\left(\frac{\partial A_l}{\partial x^k}\right)} = -\frac{1}{4\pi c} F^{kl}$; therefore,

$$T_i^k = -\frac{1}{4\pi c} \frac{\partial A_l}{\partial x^i} F^{kl} + \frac{1}{16\pi c} \delta_i^k F_{lm} F^{lm}.$$

Now we raise the index $i$ by means of the Minkowski metric $g^{ik}$ to obtain

$$T^{ik} = -\frac{g^{im}}{4\pi c} \frac{\partial A_l}{\partial x^m} F^{kl} + \frac{1}{16\pi c} g^{ik} F_{lm} F^{lm}.$$

This tensor is not symmetric, but we can symmetrize it by subtracting the sum $\frac{1}{4\pi c} \frac{\partial A^i}{\partial x^l} F^{kl}$. This sum is representable as $\frac{\partial}{\partial x^l}(\psi^{ikl})$:

$$\frac{\partial A^i}{\partial x^l} F^{kl} = \frac{\partial}{\partial x^l}(A^i F^{kl}) - A^i \frac{\partial F^{kl}}{\partial x^l} = \frac{\partial}{\partial x^l}(A^i F^{kl}),$$

since $\frac{\partial F^{kl}}{\partial x^l} = 0$ by virtue of the Maxwell equations describing the field in the absence of charges ($j = 0$). As shown in Lemma 14.1, the quantity $\frac{\partial}{\partial x^l}(\psi^{ikl})$ may be added to the energy-momentum tensor without affecting the momentum vector of the system.

Since $\frac{\partial A_l}{\partial x^i} - \frac{\partial A_i}{\partial x^l} = F_{il}$, we finally obtain the symmetric energy-momentum tensor in the form

$$T^{ik} = \frac{1}{4\pi c}\left(-F^{il} F_l^k + \frac{1}{4} g^{ik} F_{lm} F^{lm}\right). \tag{14.16}$$

**14.2.3. Einstein equations. Hilbert functional.** Consider a four-dimensional manifold $M^4$ endowed with a pseudo-Riemannian metric $g_{ik}$ of signature $(+---)$. At each point $x_0 \in M^4$ the metric can be reduced by a change of coordinates to the Minkowski metric

$$g_{ik} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Therefore, $g = \det g_{ik} < 0$ and $|g| = -g$.

Let $d\mu = \sqrt{-g}\,d^4x$ be the volume form on $M^n$, let $\Gamma^i_{jk}$ be the symmetric connection compatible with the metric, and let $R^i_{jkl}$ be the Riemann curvature tensor corresponding to this connection. Recall that the *Ricci tensor* $R_{ik} = R^q_{iqk} = g^{lm}R_{limk}$ is written as

$$R_{ik} = \frac{\partial \Gamma^l_{ik}}{\partial x^l} - \frac{\partial \Gamma^l_{il}}{\partial x^k} + \Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{mk}$$

and the *scalar curvature* $R$ equals

$$R = g^{ik}R_{ik}.$$

According to the principles of general relativity, the gravitational field in the space-time $M^4$ is the metric $g_{ij}$. The Einstein equations for the gravitational field (in empty space) are obtained as the Euler–Lagrange equations under variation of the action of this field. The action functional was introduced by Hilbert, and (up to multiplication by a physical constant) it is

$$S_g = \int R\,d\mu = \int R\sqrt{-g}\,d^4x.$$

When $M^4$ is four-dimensional space with coordinates $x^0, \dots, x^3$, the integral is taken over a domain $D$ bounded by two space-like hypersurfaces. These hypersurfaces are analogs of the time level surfaces for $t = t_1, t_2$ in the Minkowski space.

We will find the Euler–Lagrange equations for the most general functional

$$S_g = \int_M R\sqrt{|g|}\,d^nx,$$

where $M$ is an $n$-dimensional manifold with metric $g_{ij}$. Variation of this action implies variation of the components of the metric $g_{ij}$.

**Theorem 14.6.** *The following identity holds:*

$$\frac{\delta \int R\sqrt{|g|}\,d^nx}{\delta g^{ij}} = \left(R_{ij} - \frac{1}{2}\,Rg_{ij}\right)\sqrt{|g|},$$

*from which*

$$\delta S_g = \int \left(R_{ij} - \frac{1}{2}\,Rg_{ij}\right)\delta g^{ij}\sqrt{|g|}\,d^nx.$$

**Proof.** We have

$$\begin{aligned}\delta \int R\sqrt{|g|}\,d^nx &= \delta \int g^{ik}R_{ik}\sqrt{|g|}\,d^nx \\ &= \int (R_{ik}\sqrt{|g|}\delta g^{ik} + R_{ik}g^{ik}\delta(\sqrt{|g|}) + g^{ik}\sqrt{|g|}\delta R_{ik})\,d^nx.\end{aligned}$$

First we will find $\delta(\sqrt{|g|})$. Let $\Delta^{ik}$ be the cofactor of the element $g_{ik}$ in the matrix $(g_{lm})$. For the determinant of this matrix, we write down the

expansion by minors of the $k$th column: $g = \sum_i g_{ik}\Delta^{ik}$. Clearly, $\delta g = (\delta g_{ik})\Delta^{ik}$ with summation now over $i$ and $k$, since the differential $\delta g_{ik}$ of each component $g_{ik}$ must be multiplied (when collecting the like terms) by the coefficient of this component in the expression for $g$, i.e., by $\Delta^{ik}$. Since $g^{ik} = \frac{\Delta^{ik}}{g}$, we have $\Delta^{ik} = gg^{ik}$, which implies $\delta g = gg^{ik}\,\delta g_{ik}$. Note that $g^{ik}g_{ik} = \delta^i_i = n = \dim M$; therefore, $(\delta g^{ik})g_{ik} + g^{ik}(\delta g_{ik}) = 0$, which implies the equality $\delta g = -gg_{ik}\,\delta g^{ik}$. Assume for definiteness that $|g| = -g$. We have

$$\delta(\sqrt{-g}) = \frac{-1}{2\sqrt{-g}}\,\delta g = \frac{1}{2\sqrt{-g}}\,g \cdot g_{ik}\delta g^{ik} = -\frac{1}{2}\,\sqrt{-g}g_{ik}\delta g^{ik}.$$

By similar calculations for $|g| = g$ we arrive at the general formula

$$\delta(\sqrt{|g|}) = -\frac{1}{2}\,\sqrt{|g|}g_{ik}\delta g^{ik}.$$

Hence we obtain

$$\delta\int R\sqrt{|g|}\,d^nx = \int\Big(R_{ik} - \frac{1}{2}\,Rg_{ik}\Big)\delta g^{ik}\sqrt{|g|}\,d^nx + \int g^{ik}(\delta R_{ik})\sqrt{|g|}\,d^nx.$$

Now we will find $\delta R_{ik}$. Recall that by Lemma 10.1 the variations $\delta\Gamma^i_{jk}$ form a tensor, although the Christoffel symbols $\Gamma^i_{jk}$ do not.

To calculate $\delta R_{ik}$, fix an arbitrary point and introduce the geodesic coordinate system in a neighborhood of this point (see 10.3.1). This means that $\Gamma^k_{ij} = 0$ at this point, since $\frac{\partial}{\partial x^\alpha}\,(g^{ik}) = 0$ at the selected point. We have

$$\begin{aligned} g^{ik}\delta R_{ik} &= \delta\Big(\frac{\partial\Gamma^l_{ik}}{\partial x^l} - \frac{\partial\Gamma^l_{il}}{\partial x^k} + \Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{km}\Big)g^{ik} \\ &= g^{ik}\Big(\frac{\partial}{\partial x^l}\,(\delta\Gamma^l_{ik}) - \frac{\partial}{\partial x^k}\,(\delta\Gamma^l_{il})\Big) = g^{ik}\frac{\partial}{\partial x^l}\,(\delta\Gamma^l_{ik}) - g^{il}\frac{\partial}{\partial x^l}\,(\delta\Gamma^k_{ik}) \\ &= \frac{\partial}{\partial x^l}\,(g^{ik}\delta\Gamma^l_{ik} - g^{il}\delta\Gamma^k_{ik}) = \frac{\partial W^l}{\partial x^l}, \\ W^l &= g^{ik}\delta\Gamma^l_{ik} - g^{il}\delta\Gamma^k_{ik}. \end{aligned}$$

Since $\delta\Gamma^i_{jk}$ is a tensor, the quantities $W^l$ also form a tensor; hence its divergence $\frac{\partial W^l}{\partial x^l}$ in our special coordinate system remains unchanged under transition to any other curvilinear coordinate system. Here we must use the invariant definition $\operatorname{div} T = \nabla_i T^i$, since $\frac{\partial}{\partial x^i}(T^i)$ is not a tensor relative arbitrary changes of coordinates. Recall (see Section 10.3) the explicit formula for $\operatorname{div} T$ in an arbitrary coordinate system, relative to a symmetric connection compatible with metric:

$$\nabla_i T^i = \frac{\partial T^i}{\partial x^i} + T^l\frac{\partial}{\partial x^l}\,(\log\sqrt{|g|}) = \frac{1}{\sqrt{|g|}}\,\frac{\partial}{\partial x^l}\,(\sqrt{|g|}T^l).$$

We obtain $g^{ik}\delta R_{ik} = \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^l}(\sqrt{|g|}W^l)$. Thus we transformed the integral $\int g^{ik}\delta R_{ik}\sqrt{|g|}\,d^n x$ to $\int \frac{\partial}{\partial x^l}(\sqrt{|g|}W^l)\,d^n x$. By the Stokes formula, this integral is equal to the integral of $|g|W^l$ over the boundary $\partial D$ of the domain. Since the variation of the field vanishes on $\partial D$, this integral is also equal to zero. Thus we obtain

$$\delta S_g = \int \Big(R_{ik} - \frac{1}{2}\,Rg_{ik}\Big)\delta g^{ik}\sqrt{|g|}\,d^n x.$$

The proof of Theorem 14.6 is completed. □

In general relativity the action consists of two terms: $S = S_g + S_m$, where $S_m$ is the material action. The Euler–Lagrange equations for it become

$$R_{ik} - \frac{1}{2}\,Rg_{ik} = -\frac{\delta S_m}{\delta g^{ik}}. \tag{14.17}$$

The right-hand side is equal, up to multiplication by a physical constant, to the energy-momentum tensor $T_{ik}$ of the system. For example, in the case of electromagnetic field we obtain the symmetric energy-momentum tensor. We will discuss this in more detail in Section 15.1.

Note that equations (14.17) are not linear and the sum of two solutions is not necessarily a solution.

For $\dim M = n \neq 2$ the equations

$$R_{ik} - \frac{R}{2}\,g_{ik} = 0$$

(which in relativity theory are the equations of gravitational field in the empty space, i.e., for $T_{ik} = 0$) become

$$R_{ik} = 0$$

(the metric is *Ricci flat*). Indeed, we have

$$g^{ik}\Big(R_{ik} - \frac{1}{2}\,Rg_{ik}\Big) = R - \frac{n}{2}\,R = 0. \tag{14.18}$$

It does not follow from $R_{ik} = 0$ that the empty space-time is flat: the condition that the Ricci tensor vanishes is not sufficient for that. The space is flat if the entire Riemann curvature tensor $R^i_{jkl}$ vanishes. In the three-dimensional space the Riemann tensor can be expressed in terms of the Ricci tensor (see 10.2.3), so that the condition $R_{ik} = 0$ implies that the entire space is flat.

In the two-dimensional case we always have $R_{ik} = \frac{1}{2}\,Rg_{ik}$, hence the integral $S_g$ remains unchanged under variations of the metric, and the scalar curvature $R$ equals twice the Gaussian curvature $K$. This entails the following theorem.

**Theorem 14.7.** *The quantity* $S[g] = \int K\sqrt{g}\,dx^1 \wedge dx^2$, *where* $K$ *is the Gaussian curvature, does not change under local variations of the metric* $g_{ij}$, $i, j = 1, 2$.

If the surface imbedded into $\mathbb{R}^3$ is closed, then the fact that variations of the metric are local is immaterial, and we obtain the following result.

**Corollary 14.1** (Gauss–Bonnet theorem). *The integral of the Gaussian curvature over a closed oriented surface in the three-dimensional Euclidean space remains unchanged under smooth deformations of the surface:*

$$\int K\sqrt{g}\,dx^1 \wedge dx^2 = \text{const}.$$

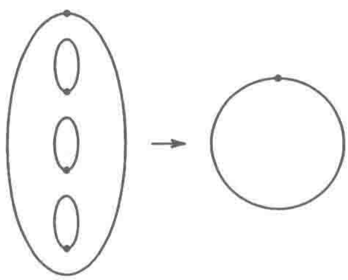

**Figure 14.2**

It can be shown that this integral is equal to

$$\int K\sqrt{g}\,dx^1 \wedge dx^2 = 2\pi(2 - 2g),$$

where $\chi = 2 - 2g$ is the Euler characteristic of the surface which is a sphere with $g$ handles.

Indeed, consider the Gauss map for the surface shown in Figure 14.2. We see that the degree of this map equals $(1 - g)$. The surface area of the unit sphere is $4\pi$, and by Theorem 12.7 the integral of the Gaussian curvature is equal, up to $4\pi$, to the degree of the Gauss map.

As a result, for a sphere with $g$ handles we have

$$\int K\sqrt{g}\,dx^1 \wedge dx^2 = 4\pi(1 - g).$$

Multiplication of the metric $g_{ij}$ by a positive constant $\mu^2$, $g_{ij} \to \mu^2 g_{ij}$, gives rise to the following transformations:

$$R \to \frac{1}{\mu^2} R, \quad |g| \to \mu^n |g|, \quad S_g \to \mu^{n-2} S_g.$$

This means that for $n \neq 2$ the functional $S_g$ is homogeneous. If it is not equal to zero, its value may be reduced simply by rescaling: $g_{ij} \to \mu^2 g_{ij}$.

Therefore, for $n \geq 3$, it makes no sense to speak about critical points of the functional $S_g$, but it is reasonable to consider problems on conditional extrema and find critical points of $S_g$ for metrics of fixed (e.g., unit) volume. By the Lagrange rule we must find critical points of the functional

$$S_{g,\Lambda} = \int (R + 2\Lambda)\sqrt{|g|}\, d^n x.$$

Variation of this functional in components of the metric yields the equations

$$R_{ik} - \frac{1}{2} R g_{ik} = \Lambda g_{ik}, \quad \Lambda = \text{const}.$$

The (Riemannian or pseudo-Riemannian) metrics satisfying these equations are called the *Einstein metrics*. One can prove similarly to (14.18) that for $n \geq 3$ the Einstein metrics have constant scalar curvature. For $n = 3$ the Riemann tensor can be recovered from the Ricci tensor, and a Riemannian metric is an Einstein metric if and only if it has a constant sectional curvature.

**14.2.4. Harmonic functions and the Hodge expansion.** Let $M^n$ be an oriented smooth manifold with Riemannian metric, and let

$$C^*(M^n) = \sum_{k=0}^{n} C^k(M^n)$$

be the algebra of smooth differential forms on $M^n$ graded by the degrees, $\deg = k$, of the forms.

We specify a bilinear scalar product on the space $C^*(M^n)$ by the following rules:

1) if the degrees of the forms $\omega_1$ and $\omega_2$ are different, then $\langle \omega_1, \omega_2 \rangle = 0$;

2) if the forms $\omega_1$ and $\omega_2$ have the same degree, then

$$\langle \omega_1, \omega_2 \rangle = \int_{M^n} \omega_1 \wedge *\omega_2. \tag{14.19}$$

For convergence of the integral we must require that the manifold be compact or the forms be compactly supported.

We will assume that $M^n$ is a compact manifold without boundary.

Any metric at a given point $P$ may be reduced to the Euclidean form by a change of coordinates:

$$ds^2 = \delta_{ij}\, dx^i\, dx^j.$$

If the coordinates $(x^1, \ldots, x^n)$ specify the positive orientation, then

$$*\omega = *(f(x)\, dx^1 \wedge \cdots \wedge dx^k) = f(x)\, dx^{k+1} \wedge \cdots \wedge dx^n,$$
$$\omega \wedge *\omega = |f|^2 \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^n$$

at this point (in these coordinates $g = 1$ at the point $P$). Therefore, this scalar product is positive definite,

$$\langle \omega, \omega \rangle \geq 0$$

(with equality only for $\omega = 0$), and symmetric,

$$\langle \omega_1, \omega_2 \rangle = \langle \omega_2, \omega_1 \rangle.$$

We have the following obvious lemma.

**Lemma 14.2.** *The exterior differential $d$ and the divergence operator $\delta$ are conjugate relative to the scalar product* (14.19):

$$d^* = \delta.$$

Recall that the action of $\delta$ on forms of degree $k$ is defined as

$$\delta = (-1)^k *^{-1} d * = (-1)^{nk+n+1} * d *.$$

**Proof.** Let $\omega_1$ and $\omega_2$ be differential forms of degrees $k-1$ and $k$, respectively. Then, by a direct calculation, we obtain

$$\langle d\omega_1, \omega_2 \rangle = \int d\omega_1 \wedge *\omega_2 = \int d(\omega_1 \wedge *\omega_2) + (-1)^k \int \omega_1 \wedge d{*}\omega_2.$$

Since the integral of the exterior differential of a form over $M^n$ equals zero by the Stokes theorem, and

$$*^2 = (-1)^{(n-k)k}$$

on the space of $k$-forms, we have

$$(-1)^k \int \omega_1 \wedge d{*}\,\omega_2 = \int \omega_1 \wedge * \left((-1)^k *^{-1} d\, *\right)\omega_2 = \langle \omega_1, \delta\omega_2 \rangle.$$

Hence the lemma. □

The Hodge functional is defined as

$$H(\omega) = \langle (d + d^*)\omega, (d + d^*)\omega \rangle.$$

Since $d^2 = 0$ and $(d^*)^2 = \delta^2 = 0$ (see 9.1.2), we have $\langle d\omega, d^*\omega \rangle = \langle d^2\omega_1, \omega_2 \rangle = 0$, and the Hodge functional becomes

$$H(\omega) = |d\omega|^2 + |d^*\omega|^2.$$

The critical points of the Hodge functional are referred to as *harmonic forms.*

**Theorem 14.8.** *A form $\omega$ is harmonic if and only if it satisfies the equation*

$$\Delta\omega = (dd^* + d^*d)\omega = 0. \tag{14.20}$$

**Proof.** We find the variational derivative of the Hodge functional. Since the operator $A = d + d^*$ is selfadjoint, for any variation $\omega \to \omega + t\alpha$ we have

$$\begin{aligned}\langle A(\omega + t\alpha), A(\omega + t\alpha)\rangle &= |A\omega|^2 + 2t\langle A\omega, A\alpha\rangle + O(t^2)\\ &= |A\omega|^2 + 2t\langle A^*A\omega, \alpha\rangle + O(t^2) = |A\omega|^2 + 2t\langle A^2\omega, \alpha\rangle + O(t^2),\end{aligned}$$

so that the variational derivative is

$$\frac{\delta H}{\delta \omega} = (d + d^*)^2\omega.$$

At the same time $(d + d^*)^2 = dd^* + d^*d$. Hence the theorem. □

The operator

$$\Delta = dd^* + d^*d = d\delta + \delta d$$

is called the *Laplace operator*. It takes $k$-forms into $k$-forms and acts on the coefficients of these forms as a second-order differential operator. When applied to functions, this operator coincides with the Laplace–Beltrami operator.

Obviously, the Laplace operator is selfadjoint:

$$\Delta^* = \Delta,$$

and so is the operator $D = d + d^* = d + \delta$.

Harmonic forms are obtained as solutions of the second-order equation (14.20). At the same time, as was shown by Hodge, they can in fact be found from a system of two first-order equations.

**Theorem 14.9.** *A form $\omega$ is harmonic if and only if it satisfies the equations*

$$d\omega = 0, \quad d^*\omega = \delta\omega = 0. \tag{14.21}$$

**Proof.** It is obvious that relations (14.21) imply (14.20). Suppose now that $(dd^* + d^*d)\omega = 0$. Then

$$\langle (dd^* + d^*d)\omega, \omega\rangle = \langle d^*\omega, d^*\omega\rangle + \langle d\omega, d\omega\rangle = |d^*\omega|^2 + |d\omega|^2 = 0,$$

which implies equalities (14.21). Hence the theorem. □

Now we introduce the following subspaces in $C^*(M^n)$:

1) $B_d = B^*_d(M^n)$, the space of all exact forms, i.e., all the forms written as $d\alpha$;

2) $B_\delta = B^*_\delta(M^n)$, the space of all forms written as $\delta\beta$;

3) $\mathcal{H} = \mathcal{H}^*(M^n)$, the space of all harmonic forms.

**Lemma 14.3.** 1. *The spaces $B_d$, $B_\delta$, and $\mathcal{H}$ are pairwise orthogonal.*

2. *If a form $\omega$ is orthogonal to the spaces $B_d$ and $B_\delta$, then $\omega$ is a harmonic form.*

**Proof.** Both statements are verified directly. We have

$$\langle d\alpha, d^*\beta\rangle = \langle d^2\alpha, \beta\rangle = 0,$$

since $d^2 = 0$. Hence the subspaces $B_d$ and $B_\delta$ are orthogonal.

Let $d\omega = d^*\omega = 0$ (i.e., the form $\omega$ is harmonic). Then

$$\begin{aligned} \langle \omega, d\alpha\rangle &= \langle d^*\omega, \alpha\rangle = 0, \\ \langle \omega, d^*\beta\rangle &= \langle d\omega, \beta\rangle = 0. \end{aligned} \tag{14.22}$$

Therefore, the space $\mathcal{H}$ is orthogonal to $B_d$ and $B_\delta$. This proves assertion 1.

Conversely, if equalities (14.22) hold for all forms $\alpha$ and $\beta$, then $d\omega = d^*\omega = 0$. This proves assertion 2 and hence the lemma. □

The following fact to be stated without proof is established by means of the theory of elliptic equations.

**Theorem 14.10** (Hodge decomposition). *Each smooth form $\omega$ on a closed manifold is uniquely representable as the sum of smooth forms,*

$$\omega = d\alpha + d^*\beta + \varphi,$$

*where $\varphi$ is a harmonic form.*

Note that a form is closed if and only if it can be written as $d\alpha + \varphi$ (i.e., the term $d^*\beta$ in the Hodge decomposition vanishes). Indeed, $d^2\alpha = d\varphi = 0$, and if $dd^*\beta = 0$, then

$$\langle dd^*\beta, \beta\rangle = \langle d^*\beta, d^*\beta\rangle = 0.$$

Therefore, the space of closed forms is $Z = B_d + \mathcal{H}$, and the de Rham cohomology group $H^* = Z/B_d$ (see 9.3.1) is isomorphic to the space of harmonic forms. Hence we obtain the following result.

**Corollary 14.2.** *Each cohomology class $[\omega] \in H^*(M^n; \mathbb{R})$ of a closed manifold $M^n$ is uniquely represented by a harmonic form $\omega$.*

Note that the linear isomorphism $\mathcal{H} = H^*$ is not extendable to an isomorphism of rings, since the exterior product of a pair of harmonic forms is not necessarily a harmonic form.

EXAMPLE. FLAT TORI. Let $M^n = T^n = \mathbb{R}^n/\Lambda$ be a torus with Euclidean metric $\delta_{ij}\, dx^i\, dx^j$ (here $(x^1, \dots, x^n)$ are linear coordinates on the torus determined up to vectors of the lattice $\Lambda$). Then the operator $\Delta$ in these coordinates coincides up to sign with the usual Laplace operator acting on the coefficients of the forms. For brevity, we will demonstrate this only in the two-dimensional case.

Let $n = 2$ and $\omega = a\,dx + b\,dy$. The operator $\delta$ acts on 1-forms in any dimension as $\delta = -{*d*}$. The subscripts $x$ and $y$ will denote differentiation with respect to these variables. We have

$$\begin{aligned}
a\,dx + b\,dy &\xrightarrow{*} -b\,dx + a\,dy \xrightarrow{d} (b_y + a_x)\,dx \wedge dy \\
&\xrightarrow{-*} -(a_x + b_y) \xrightarrow{d} dd^*\omega = -(a_{xx} + b_{xy})\,dx - (a_{xy} + b_{yy})\,dy, \\
a\,dx + b\,dy &\xrightarrow{d} (-a_y + b_x)\,dx \wedge dy \xrightarrow{*} (b_x - a_y) \\
&\xrightarrow{d} (b_{xx} - a_{xy})\,dx + (b_{xy} - a_{yy})\,dy \\
&\xrightarrow{-*} d^*d\omega = -(a_{yy} - b_{xy})\,dx - (-a_{xy} + b_{xx})\,dy.
\end{aligned}$$

Thus we obtain

$$\Delta(a\,dx + b\,dy) = -(a_{xx} + a_{yy})\,dx - (b_{xx} + b_{yy})\,dy.$$

It remains to note that by the maximum principle for harmonic functions the doubly periodic functions $a$ and $b$ bounded in the entire space $\mathbb{R}^2$ are constant. Hence we obtain the following result.

**Corollary 14.3.** *Harmonic forms on a torus are precisely the forms with constant coefficients (in Euclidean coordinates).*

The ring of forms with constant coefficients on the torus is isomorphic to its de Rham cohomology ring. In fact, we have already proved this when presenting the averaging method for cohomology computations in homogeneous spaces (see 9.3.3).

Let $T^n = \mathbb{R}^n/\mathbb{Z}^n$ be a flat torus with cubic lattice of periods. In Euclidean coordinates we do not distinguish between vector and covector fields, hence we may speak about the Hodge decomposition of vector fields. It is very simple.

Let $v = \sum_{k\in\mathbb{Z}^n} a_k e^{2\pi i\langle k,x\rangle}$ be the Fourier expansion of the vector field $v$. Decompose each vector coefficient $a_k$ for $k \neq 0$ into the sum $a_k = a'_k + a''_k$, where the vectors $a'$ and $k$ are proportional and the vectors $a''_k$ and $k$ are orthogonal, $\langle a''_k, k\rangle = 0$. Set

$$v_0 = a_0, \quad v' = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} a'_k e^{2\pi i\langle k,x\rangle}, \quad v'' = \sum_{k\in\mathbb{Z}^n\setminus\{0\}} a''_k e^{2\pi i\langle k,x\rangle}.$$

We have the following decomposition:

$$\begin{gathered}
v = v_0 + v' + v'', \\
v' = d\alpha, \quad v'' = \delta\beta, \quad dv_0 = \delta v_0 = 0,
\end{gathered}$$

where the field $v_0$ has constant coefficients (is harmonic), the field $v'$ is gradient: $v' = d\alpha$, and the field $v''$ is divergence-free: $\delta v'' = 0$ (which means

that $v'' = \delta\beta$). This decomposition was first obtained in elasticity theory and is called the *Helmholtz decomposition.*

Consider now a bounded domain $U$ in three-dimensional Euclidean space. Obviously, we may assume that this domain lies on a flat torus and all the forms on this domain can be extended to smooth forms. Then the following assertion holds:

*Each vector field $v$ on the domain $U$ is decomposed into the sum of a gradient and divergence-free vector fields.*

This decomposition is not unique: each term is defined up to addition of a field with constant coefficients.

**14.2.5. The Dirichlet functional and harmonic mappings.** Let $M^n$ and $N^q$ be two Riemann manifolds with metrics $g_{ij}$ and $h_{\alpha\beta}$, respectively. We define the *Dirichlet functional* on the space of smooth mappings from $M^n$ to $N^q$ by the following formula:

$$S(f) = \frac{1}{2}\int_M g^{ij} h_{\alpha\beta}\,\frac{\partial f^\alpha}{\partial x^i}\,\frac{\partial f^\beta}{\partial x^j}\,\sqrt{g}\,dx^1\wedge\cdots\wedge dx^n, \tag{14.23}$$

where $f\colon M^n \to N^q$ is a smooth mapping.

A mapping $f\colon M^n \to N^q$ will be called *harmonic* if it is a critical point of the functional $S$ of the form (14.23).

**Theorem 14.11.** *A mapping $f\colon M^n \to N^q$ is harmonic if and only if it satisfies the equation*

$$\Delta f^\gamma + g^{ij}\tilde\Gamma^\gamma_{\alpha\beta}\,\frac{\partial f^\alpha}{\partial x^i}\,\frac{\partial f^\beta}{\partial x^j} = 0 \tag{14.24}$$

*for all $\gamma = 1,\dots,q$, where $\Delta$ is the Laplace–Beltrami operator on $M^n$:*

$$\Delta f = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\Bigl(g^{ij}\sqrt{g}\,\frac{\partial}{\partial x^j}\Bigr)$$

*and $\tilde\Gamma^\gamma_{\alpha\beta}$, $\alpha,\beta,\gamma = 1,\dots,q$, are the Christoffel symbols of the unique symmetric connection on $N^q$ that agrees with the metric $h_{\alpha\beta}$.*

**Proof.** Consider a $t$-parametrized perturbation $f_t$ of the mapping $f = f_0$ and assume that this perturbation has a compact support disjoint from the boundary of $M^n$. Set $v = \frac{df_t}{dt}\big|_{t=0}$. We have

$$\begin{aligned}\frac{dS(f_t)}{dt}\Big|_{t=0} = \frac{1}{2}\int_{M^n}\Big[&g^{ij}\,\frac{\partial h_{\alpha\beta}}{\partial\varphi^\lambda}\,v^\lambda\,\frac{\partial f^\alpha}{\partial x^i}\,\frac{\partial f^\beta}{\partial x^j}\,\sqrt{g}\\ &+ g^{ij}h_{\alpha\beta}\,\frac{\partial v^\alpha}{\partial x^i}\,\frac{\partial f^\beta}{\partial x^j}\,\sqrt{g} + g^{ij}h_{\alpha\beta}\,\frac{\partial f^\alpha}{\partial x^i}\,\frac{\partial v^\beta}{\partial x^j}\,\sqrt{g}\Big]\,dx^1\wedge\cdots\wedge dx^n,\end{aligned}$$

and, integrating by parts, we obtain

$$\frac{dS(f_t)}{dt}\Big|_{t=0} = \frac{1}{2}\int_{M^n}\Big[-2v^\lambda h_{\gamma\lambda}\frac{\partial}{\partial x^i}\Big(g^{ij}\sqrt{g}\,\frac{\partial f^\gamma}{\partial x^j}\Big) + v^\lambda g^{ij}\sqrt{g}\,\frac{\partial f^\alpha}{\partial x^i}\frac{\partial f^\beta}{\partial x^j}\Big(\frac{\partial h_{\alpha\beta}}{\partial\varphi^\lambda} - \frac{\partial h_{\lambda\beta}}{\partial\varphi^\alpha} - \frac{\partial h_{\alpha\lambda}}{\partial\varphi^\beta}\Big)\Big]\,dx^1\wedge\cdots\wedge dx^n.$$

Since

$$\frac{1}{2}\Big(\frac{\partial h_{\alpha\lambda}}{\partial\varphi^\beta} + \frac{\partial h_{\lambda\beta}}{\partial\varphi^\alpha} - \frac{\partial h_{\alpha\beta}}{\partial\varphi^\lambda}\Big) = \tilde{\Gamma}^\gamma_{\alpha\beta}h_{\lambda\gamma},$$

we obtain as a final result that

$$\frac{dS(f_t)}{dt}\Big|_{t=0} = -\int_{M^n} h_{\gamma\lambda}v^\lambda\tau^\gamma(f)\sqrt{g}\,dx^1\wedge\cdots\wedge dx^n,$$

where $\tau(f)$ is the tension tensor,

$$\tau^\gamma(f) = \Delta f^\gamma + g^{ij}\tilde{\Gamma}^\gamma_{\alpha\beta}\frac{\partial f^\alpha}{\partial x^i}\frac{\partial f^\beta}{\partial x^j}.$$

Obviously, the mapping $f$ is harmonic if and only if $\tau(f) = 0$. Hence the theorem. ☐

EXAMPLE. GEODESICS. Let $M^1$ be a straight line or an interval on the line. Any metric can be reduced to the form $dt^2$ ($g_{11}\equiv 1$) by a change of coordinates. Then the equation (14.24) for the mapping $f(t) = (y^1(t),\dots,y^q(t))$ becomes the equation for geodesics in $N^q$:

$$\ddot{y}^i + \tilde{\Gamma}^i_{jk}\dot{y}^j\dot{y}^k = 0, \quad i = 1,\dots,q.$$

Therefore, the geodesics, and only they, are one-dimensional harmonic mappings.

Note that the harmonicity property of a multidimensional mapping depends on the choice of metrics on both manifolds $M^n$ and $N^q$. In the two-dimensional case, the Dirichlet functional $S$ has the following property of conformal invariance.

**Lemma 14.4.** *The value of the Dirichlet functional $S(f)$ for a mapping of a two-dimensional manifold $f\colon M^2\to N^q$ remains unchanged if the metric $g_{ij}$ on the manifold $M^2$ is replaced by a conformally invariant metric $g'_{ij}(x) = \lambda(x)g_{ij}(x)$.*

The lemma follows from the fact that the metric $g_{ij}$ occurs in the integrand of the Dirichlet functional via the expression $g^{ij}\sqrt{g}$, which is conformally invariant.

**Corollary 14.4.** *If a mapping $f\colon M^2\to N^q$ of a two-dimensional manifold $M^2$ with metric $g_{ij}$ is harmonic, then it is harmonic for any conformally equivalent metric $\lambda g_{ij}$.*

The Dirichlet functional is a natural extension of the volume functional. Indeed, suppose the metric $g_{ij}$ is induced by a metric $h_{\alpha\beta}$ under an immersion of $M^n$ into $N^q$. At each point of $M^n$ the vectors $f_i = \left(\frac{\partial f^\alpha}{\partial x^i}\right)$, $i = 1, \ldots, n$, determine a basis of tangent vectors and

$$h_{\alpha\beta}\frac{\partial f^\alpha}{\partial x^i}\frac{\partial f^\beta}{\partial x^j} = \langle f_i, f_j\rangle = g_{ij}.$$

Moreover, the integrand of the Dirichlet functional equals

$$g^{ij}g_{ij}\sqrt{g} = n\sqrt{g}.$$

Hence we obtain the following result.

**Lemma 14.5.** *If the metric $g_{ij}$ on $M^n$ is induced by a metric $h_{\alpha\beta}$ on $N^q$ under an immersion $f\colon M^n \to N^q$, then the Dirichlet functional is proportional to the volume of the manifold $M^n$:*

$$S(f) = \frac{n}{2}\int_{M^n}\sqrt{g}\,dx^1\wedge\cdots\wedge dx^n.$$

The critical points of the volume functional are the *minimal submanifolds* (we already considered them in the two-dimensional case with $N^q = \mathbb{R}^3$ when dealing with minimal surfaces; see 14.2.1).

Before deriving the equation for minimal submanifolds, we must introduce necessary notions.

For any submanifold $M^n \subset N^q$ with induced metric, denote by $x^1, \ldots, x^n$ local coordinates in $M^n$, and by $x^{n+1}, \ldots, x^q$, the coordinates in $N^q$ normal to $M^n$.

Since in these coordinates $h_{ij} = g_{ij}$ for $i, j = 1, \ldots, n$, the Christoffel symbols $\Gamma^j_{kl}$ and $\tilde{\Gamma}^j_{kl}$ in $M^n$ and $N^q$, respectively, coincide for $1 \le j, k, l \le n$:

$$\Gamma^j_{kl} = \tilde{\Gamma}^j_{kl} \quad \text{for} \quad 1 \le j, k, l \le n.$$

This follows from the formula

$$\tilde{\Gamma}^j_{kl} = \frac{1}{2}g^{jm}\Big(\frac{\partial h_{km}}{\partial x^l} + \frac{\partial h_{lm}}{\partial x^k} - \frac{\partial h_{kl}}{\partial x^m}\Big)$$

for Christoffel symbols.

The *second fundamental form $B(\xi, \eta)$ of a submanifold* is a bilinear form in vectors $\xi, \eta$ tangent to $M^n$ at some point $P$, and its value is a vector normal to $M^n$ at the same point $P$.

To define it, we extend $\xi$ and $\eta$ to vector fields in a neighborhood of the point $P$, take the covariant derivative of the field $\nabla_\eta\xi$ relative to the connection on $N^q$, and set $B(\xi, \eta)$ to be the orthogonal projection of the

vector $\nabla_\eta \xi$ to the subspace orthogonal to $M^n$ in the tangent space to $N^q$ at the point $P$:

$$\nabla_\eta \xi = B(\xi,\eta) + v, \quad v \in TM^n, \ B(\xi,\eta) \perp TM^n.$$

It turns out that the vector $B(\xi,\eta)$ depends only on the vectors $\xi$ and $\eta$ at the point $P$.

For $n = 2$ and $N^q = \mathbb{R}^3$ this form turns into the second fundamental form of the surface. Since in the general case

$$B^\gamma_{jk} = \frac{\partial^2 f^\gamma}{\partial x^j \partial x^k} + \frac{\partial f^\alpha}{\partial x^j}\frac{\partial f^\beta}{\partial x^k}\Gamma'^\gamma_{\alpha\beta} - \Gamma^l_{jk}\frac{\partial f^\gamma}{\partial x^l}, \quad j,k = 1,\dots,n,$$

it is obvious that in the coordinates chosen above the following analogs of derivational equations hold:

$$B^\gamma_{kl} = \tilde{\Gamma}^\gamma_{kl} - \Gamma^\gamma_{kl} = 0 \quad \text{for } \gamma \le n, \quad B^\gamma_{kl} = \tilde{\Gamma}^\gamma_{kl} \quad \text{for } \gamma > n.$$

**Theorem 14.12.** *A submanifold $M^n \subset N^q$ is minimal if and only if the trace of the second fundamental form is identically equal to zero:*

$$\operatorname{Tr} B = g^{jk} B^\gamma_{jk} = 0.$$

**Proof.** We consider a variation $f_t$ of the imbedding $f : M^n \to N^q$. In contrast with Theorem 14.11, the metrics on $M^n$ are induced by the immersions $f_t$ and also vary. The volume $V = \int \sqrt{g}\,dx^1 \wedge \dots \wedge dx^n$ evolves as follows:

$$\begin{aligned}
\frac{dV}{dt} &= \int_{M^n} \frac{d\sqrt{g}}{dt} dx^1 \wedge \dots \wedge dx^n \\
&= \int_{M^n} \frac{\partial\sqrt{g}}{\partial g_{ij}}\frac{dg_{ij}}{dt} dx^1 \wedge \dots \wedge dx^n \\
&= \int_{M^n} \frac{g^{ij}\sqrt{g}}{2}\frac{dg_{ij}}{dt} dx^1 \wedge \dots \wedge dx^n \\
&= \int_{M^n} \frac{g^{ij}\sqrt{g}}{2}\frac{d}{dt}\left(h_{\alpha\beta}\frac{\partial f^\alpha}{\partial x^i}\frac{\partial f^\beta}{\partial x^j}\right) dx^1 \wedge \dots \wedge dx^n,
\end{aligned}$$

which implies that a minimal submanifold $M^n$ satisfies the harmonic equation (14.24), provided that the metric on $M^n$ is induced by the immersion.

We will use the following formula for the Laplace–Beltrami operator (which is proved by direct calculations):

$$\Delta = g^{jk}\frac{\partial^2}{\partial x^j \partial x^k} - g^{jk}\Gamma^l_{jk}\frac{\partial}{\partial x^l}.$$

After substitution of it into (14.24), the harmonic equation becomes

$$g^{jk}\Big(\frac{\partial^2 f^\gamma}{\partial x^j \partial x^k} + \frac{\partial f^\alpha}{\partial x^j}\frac{\partial f^\beta}{\partial x^k}\tilde{\Gamma}^\gamma_{\alpha\beta} - \Gamma^l_{jk}\frac{\partial f^\gamma}{\partial x^l}\Big) = 0.$$

Let $(x^1,\dots,x^q)$ be coordinates in $N^q$ such that $x^1,\dots,x^n$ are coordinates in the submanifold $M^n$ and the coordinates $x^{n+1},\dots,x^q$ are normal to $M^n$ at

the points of this submanifold. In these coordinates, the harmonic equation is

$$g^{jk} B^{\gamma}_{jk} = \operatorname{Tr} B = 0,$$

which proves the theorem. □

For arbitrary manifolds $M^n \subset N^q$, the vector

$$H = \frac{1}{2} \operatorname{Tr} B$$

is called the *mean curvature vector*. Theorem 14.12 says that minimal submanifolds are precisely the submanifolds with zero mean curvature,

$$H = 0.$$

For two-dimensional surfaces this was proved in 14.2.1.

The simplest examples of minimal submanifolds are given by *totally geodesic submanifolds*, specified by the condition $B(\xi, \eta) \equiv 0$. These are, e.g., $k$-dimensional planes in $\mathbb{R}^n$ and equator spheres $S^k$ in unit spheres $S^n$. These spheres are obtained as intersections of $S^n \subset \mathbb{R}^{n+1}$ with $(k+1)$-dimensional planes passing through $0 \in \mathbb{R}^{n+1}$.

Consider now a nontrivial example, harmonic mappings of the unit sphere $S^2$ into itself. If we delete a point from the sphere, then we can take a complex parameter $z = u + iv$ on the complement such that the metric becomes

$$dl^2 = \frac{4\, dz\, d\bar{z}}{(1+|z|^2)^2} = \frac{4(du^2 + dv^2)}{(1+u^2+v^2)^2}.$$

The sphere itself is then represented as the complex plane $\mathbb{C}$ completed with a point at infinity (see 4.2.1). This is a complex manifold that is complex-analytically diffeomorphic to the complex projective line $\mathbb{C}P^1$.

The volume form is

$$d\mu = \frac{4\, du \wedge dv}{(1+u^2+v^2)^2}, \qquad \int_{\mathbb{R}^2} d\mu = 4\pi.$$

Since in the two-dimensional case the Dirichlet functional for mappings $f\colon M^2 \to N^q$ is conformally invariant, the metric in $M^2 \setminus \{\infty\} = \mathbb{C}$ may be chosen to be Euclidean, $dl^2 = dw\, d\bar{w}$, and for $w = x + iy$, the Dirichlet functional for the mapping $z = f(w)$ from $S^2$ into $S^2$ can be written as

$$S(f) = 2 \int_{\mathbb{C}} \frac{(u_x^2 + u_y^2 + v_x^2 + v_y^2)}{(1+u^2+v^2)^2}\, dx \wedge dy.$$

By Theorem 12.7, the degree of the mapping $f$ is

$$\deg f = \frac{\int_{\mathbb{R}^2} f^*(d\mu)}{\int_{\mathbb{R}^2} d\mu} = \frac{1}{\pi} \int_{\mathbb{R}^2} \frac{(u_x v_y - u_y v_x)}{(1+u^2+v^2)^2}\, dx \wedge dy.$$

This quantity is an integer that remains unchanged under homotopies, including small perturbations of $f$.

Consider now the expression

$$S(f) - 4\pi \deg f = 2\int_{\mathbb{R}^2} \frac{(u_x - v_y)^2 + (u_y + v_x)^2}{(1+u^2+v^2)^2}\, dx \wedge dy.$$

This expression is always nonnegative, and if it vanishes for some mapping $f$, then the Dirichlet functional $S$ attains its minimum on this mapping among all mappings of degree $\deg f$.

For any degree $d \geq 0$, the equality

$$S(f) - 4\pi \deg f = 0$$

holds for mappings satisfying the Cauchy–Riemann conditions

$$u_x = v_y, \quad u_y = -v_x,$$

and only for them, i.e., for holomorphic mappings $f\colon S^2 \to S^2$ $(\partial f/\partial \bar{z} = 0)$.

Thus we have proved the following theorem.

**Theorem 14.13.** *Harmonic mappings of degree $d \geq 0$ of the unit sphere $S^2 \subset \mathbb{R}^3$ into itself are precisely the holomorphic mappings of degree $d$. The Dirichlet functional for them is equal to*

$$S = 4\pi d.$$

**14.2.6. Massive scalar and vector fields.** The examples given below are part of standard material in handbooks on field theory in courses of theoretical physics.

In physics, the action functional for a massive complex scalar field $\varphi(x)$ in the space $\mathbb{R}^{1,3}$ with Minkowski metric

$$g_{ab}\, dx^a\, dx^b = (dx^0)^2 - \sum_{\alpha=1}^{3} (dx^\alpha)^2$$

$(x^0 = ct)$ differs from the Dirichlet functional by an additional (massive) term. Namely, the action functional for such a field equals

$$S = \text{const} \cdot \int \Big[\hbar^2 \Big\langle \frac{\partial \bar{\varphi}}{\partial x}, \frac{\partial \varphi}{\partial x} \Big\rangle - m^2 c^2 \overline{\varphi(x)} \varphi(x)\Big]\, d^4x = \int \Lambda d^4 x, \tag{14.25}$$

where the bar means complex conjugation, $\hbar$ is the Planck constant (with physical dimension of action), and $c$ is the velocity of light.

The quantity $m \geq 0$ is called the *mass of the particle* described by the field $\varphi$.

The variables $\varphi$ and $\bar{\varphi}$ involved in the Lagrangian of the field,

$$\Lambda = \Lambda\Big(\varphi, \bar{\varphi}, \frac{\partial \varphi}{\partial x}\, \frac{\partial \bar{\varphi}}{\partial x}\Big),$$

are treated formally as independent variables. The Euler–Lagrange equations

$$\frac{\delta S}{\delta \varphi} = 0, \quad \frac{\delta S}{\delta \bar{\varphi}} = 0$$

for this action become the *Klein–Gordon equations*

$$(\hbar^2 \Box + m^2 c^2)\varphi = 0, \tag{14.26}$$

where

$$\Box = \frac{\partial^2}{(\partial x^0)^2} - \sum_{\alpha=1}^{3} \frac{\partial}{(\partial x^\alpha)^2}$$

is the *d'Alembert operator.*

For simplicity we set all the universal constants equal to one: $\hbar = 1$ and $c = 1$. The energy-momentum tensor has the form

$$T^{ba} = T^{ab} = g^{ac} g^{bd} \Big( \frac{\partial \bar{\varphi}}{\partial x^c} \frac{\partial \varphi}{\partial x^d} + \frac{\partial \varphi}{\partial x^c} \frac{\partial \bar{\varphi}}{\partial x^d} \Big) - g^{ab} \Lambda$$

(see 14.1.2) and the energy density equals

$$T^{00} = \sum_c \frac{\partial \bar{\varphi}}{\partial x^c} \frac{\partial \varphi}{\partial x^c} + m^2 \bar{\varphi} \varphi.$$

Since the action (14.25) is invariant under the group of transformations

$$\varphi \to e^{i\alpha} \varphi, \quad \bar{\varphi} \to e^{-i\alpha} \bar{\varphi}, \quad \alpha = \text{const},$$

this group gives rise to the conserved current

$$J^a = i g^{ab} \Big( \bar{\varphi} \frac{\partial \varphi}{\partial x^b} - \varphi \frac{\partial \bar{\varphi}}{\partial x^b} \Big). \tag{14.27}$$

The quantity

$$Q = \int_{t=\text{const}} J^0 \, d^3 x$$

is called the *charge* of the field $\varphi$.

The exterior electromagnetic field is included by the rule $p \to p + \frac{e}{c} A$, where $p = i\hbar \frac{\partial}{\partial x}$ is the momentum operator:

$$i \frac{\partial}{\partial x^a} \to i \frac{\partial}{\partial x^a} + \frac{e}{c\hbar} A_a(x).$$

Here $A_a(x)$ is the vector potential of the electromagnetic field, $e$ is the charge, and $c$ and $\hbar$ are universal constants. Taking into account the equalities $\hbar = 1$ and $c = 1$ we have

$$i \frac{\partial}{\partial x^a} \to i \frac{\partial}{\partial x^a} + e A_a(x),$$

and the total Lagrangian of the field becomes

$$\Lambda(\varphi, \bar{\varphi}, A) = \Big\langle \frac{\partial \bar{\varphi}}{\partial x^a} + ie A_a \bar{\varphi}, \frac{\partial \varphi}{\partial x^a} - ie A_a \varphi \Big\rangle - m^2 c^2 \bar{\varphi} \varphi - \frac{1}{16\pi c} F_{ab} F^{ab}, \tag{14.28}$$

where $F_{ab} = \frac{\partial A_b}{\partial x^a} - \frac{\partial A_a}{\partial x^b}$.

We leave it as a problem for the reader to verify that the action (14.28) is invariant under the following gauge transformations:

$$\varphi \to e^{i\alpha}, \quad \bar{\varphi} \to e^{-i\alpha}\bar{\varphi}, \quad A_a \to A_a + \frac{\partial \alpha}{\partial x^a}, \quad \alpha = \alpha(x).$$

Since $\bar{\varphi} \equiv \varphi$ for a real scalar field, we have

$$\Lambda = \frac{1}{2}\left\langle \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial x} \right\rangle - m^2\varphi^2$$

and the current vector (14.27) vanishes,

$$J^a = 0.$$

For this reason such a field is said to be neutral. In this case inclusion of an electromagnetic field is impossible, since the solutions of the Euler–Lagrange equations will not be real.

For the field $\mathrm{const} \cdot e^{i\langle k,x\rangle}$, which solves the free equation (14.26), the following relation holds:

$$\langle k, k\rangle = \frac{m^2c^2}{\hbar^2}.$$

In physics, such a solution is regarded as a representation of a free particle of mass $m$ with momentum $p = \hbar k$ proportional to the vector $k$. The momentum of this particle lies on the mass surface:

$$\langle p, p\rangle = m^2c^2.$$

Now we will consider massive vector fields. As before, let the fundamental constants be equal to one: $\hbar = 1$ and $c = 1$.

The Lagrangian of a complex vector field $\varphi = (\varphi_0, \varphi_1, \varphi_2, \varphi_3)$ with mass $m \neq 0$ in the space $\mathbb{R}^{1,3}$ has the form

$$\Lambda = -g^{bd}g^{ac}\frac{\partial \overline{\varphi_a}}{\partial x^b}\frac{\partial \varphi_c}{\partial x^d} + m^2 g^{ab}\overline{\varphi_a}\varphi_b,$$

with the following additional (calibration) conditions imposed on the field:

$$\frac{\partial \varphi_a}{\partial x^a} = 0, \quad \frac{\partial \overline{\varphi_a}}{\partial x^a} = 0. \tag{14.29}$$

The Euler–Lagrange equations $\frac{\partial S}{\partial \varphi} = \frac{\partial S}{\partial \bar{\varphi}} = 0$ for the action $S = \int \Lambda\, d^4x$ again have the form

$$(\Box + m^2)\varphi = 0. \tag{14.30}$$

The energy-momentum tensor is

$$T^{ab} = T^{ba} = -g^{ac}g^{bd}g^{kl}\left(\frac{\partial \overline{\varphi_k}}{\partial x^c}\frac{\partial \varphi_l}{\partial x^d} + \frac{\partial \overline{\varphi_l}}{\partial x^d}\frac{\partial \varphi_k}{\partial x^c}\right) - g^{ab}\Lambda,$$

and the group of symmetries

$$\varphi \to e^{i\alpha}\varphi, \quad \bar{\varphi} \to e^{-i\alpha}\bar{\varphi}, \quad \alpha = \mathrm{const}$$

gives rise to the conserved current

$$J^a = -ig^{ab}g^{cd}\Big(\frac{\partial \varphi_c}{\partial x^b}\,\overline{\varphi_d} - \frac{\partial \overline{\varphi_c}}{\partial x^b}\,\varphi_d\Big).$$

It can be shown that if the vector field is not subject to calibration (14.29), then in general the energy $\int T^{00}\,d^3x$ is not positive.

The electromagnetic field is again included by the rule

$$i\,\frac{\partial}{\partial x^a} \to i\,\frac{\partial}{\partial x^a} + \frac{e}{c\hbar}\,A_a.$$

The total Lagrangian taking into account the electromagnetic field is invariant under the transformations

$$\varphi \to e^{i\alpha(x)}\varphi, \quad \bar{\varphi} \to e^{-i\alpha(x)}\bar{\varphi}, \quad A_a \to A_a + \frac{\partial \alpha(x)}{\partial x^a}$$

($\hbar = c = 1$). In the case of a real vector field $\varphi$ of zero mass $m = 0$, we obtain the Lagrangian of the electromagnetic field

$$\Lambda = \mathrm{const}\cdot F_{ab}F^{ab}, \quad F_{ab} = \frac{\partial \varphi_b}{\partial x^a} - \frac{\partial \varphi_a}{\partial x^b},$$

which is invariant under the calibration transformations

$$\varphi_a \to \varphi_a + \frac{\partial \alpha(x)}{\partial x^a}.$$

The solutions to equation (14.30) of flat wave type, $\mathrm{const}\cdot e^{i\langle k,x\rangle}$, as in the case of scalar fields, possess the property

$$\langle k, k\rangle = \frac{m^2c^2}{\hbar^2},$$

and their momentum lies on the mass surface $\langle p, p\rangle = m^2c^2$.

## Exercises to Chapter 14

1. Let $T^{ij}$ be the energy-momentum tensor. Prove that

$$P^i = \frac{1}{c}\int T^{i0}\,dx^1 \wedge dx^2 \wedge dx^3$$

(the vector with components $(T^{10}/c, T^{20}/c, T^{30}/c)$ is called the *momentum density* of the system, and the quantity $T^{00}$ is the *energy density*),

$$\frac{\partial}{\partial t}\int_V T^{00}\,d^3x = -c\oint_{\partial V} T^{0\alpha}\,d\sigma_\alpha$$

(here $T^{0\alpha}\,d\sigma_\alpha = T^{01}\,dx^2 \wedge dx^3 + T^{02}\,dx^3 \wedge dx^1 + T^{03}\,dx^1 \wedge dx^2$), and

$$\frac{\partial}{\partial t}\int_V \frac{1}{c}T^{\alpha 0}\,d^3x = -\oint_{\partial V} T^{\alpha\beta}\,d\sigma_\beta.$$

The tensor $T^{\alpha\beta}$, $\alpha, \beta = 1, 2, 3$, is called the *stress tensor* (the density tensor of the momentum flux).

2. Prove that if a geodesic line on a manifold $N^q$ is tangent to a totally geodesic submanifold $M^n \subset N^q$ at some point, then this geodesic belongs entirely to $M^n$ (this property is equivalent to the property of $M^n$ being totally geodesic).

3. Prove that if a manifold $M^n \subset N^q$ is the set of fixed points of an isometric involution $F\colon M^n \to M^n$, $F^2 = 1$, then it is totally geodesic.

4. Prove that the intersections of the sphere $(x^1)^2 + \cdots + (x^n)^2 = 1$ with planes of any dimension passing through the origin are totally geodesic submanifolds of the sphere.

5. Prove that the Clifford torus defined by the equations

$$(x^1)^2 + (x^2)^2 = \frac{1}{2}, \quad (x^3)^2 + (x^4)^2 = \frac{1}{2}$$

in Euclidean space $\mathbb{R}^4$ is a minimal submanifold of the unit sphere $S^3 \subset \mathbb{R}^4$.

6. Prove that the Gauss map $G\colon M^2 \to S^2$ of a two-dimensional surface $M^2 \subset \mathbb{R}^3$ with induced metric (first fundamental form) into the unit sphere is harmonic if and only if the surface $M^2$ has a constant mean curvature, $H = \text{const}$.

7. Prove that complex submanifolds in Kähler manifolds are minimal.

8. Prove that the extremals of the functional $S(F) = \int F \wedge *F = \int F_{ik}F^{ik}\, d^4x$ subject to the condition $d(F_{ik}\, dx^i \wedge dx^k) = 0$ satisfy the Maxwell equations in empty space. Here $F_{ik}$ is a skew-symmetric tensor in the Minkowski space $\mathbb{R}^{1,3}$.

9. Consider the functional

$$S[\Gamma] = \int R\sqrt{|g|}\, dV,$$

where

$$R = g^{ik}R_{ik}, \quad R_{ik} = \frac{\partial \Gamma^l_{ik}}{\partial x^l} - \frac{\partial \Gamma^l_{il}}{\partial x^k} + \Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{mk}$$

and $g_{ik}$ is a fixed metric. Prove that the extremals $\Gamma = (\Gamma^k_{ij})$ of this functional satisfy the Christoffel formulas which express the symbols $\Gamma^k_{ij}$ in terms of the components of the metric tensor.

10. Let $M^n$ be an $n$-dimensional oriented connected compact manifold without boundary. Prove that the star operator $*$ acting on harmonic forms establishes the isomorphism

$$H^{h-k}(M^n; \mathbb{R}) = H^k(M^n; \mathbb{R}), \quad k = 0, 1, \ldots, n,$$

which is called the *Poincaré duality*.

Chapter 15

# Geometric Fields in Physics

## 15.1. Elements of Einstein's relativity theory

**15.1.1. Principles of special relativity.** Here we will state the basic notions and principles of the special relativity theory (SRT), which were partly presented in Sections 2.2, 12.3, and 14.2.

According to special relativity, created by the physicists Einstein and Lorentz and geometers Minkowski and Poincaré, the space of events is the 4-dimensional *Minkowski space* $\mathbb{R}^{1,3}$ with coordinates $x^0 = ct, x^1, x^2, x^3$ and metric

$$dl^2 = g_{ij}\, dx^i\, dx^j = (dx^0)^2 - \sum_{\alpha=1}^{3}(dx^\alpha)^2,$$

where $t$ is time and $c$ is the velocity of light in vacuum. Recall that this metric is referred to as the *Minkowski metric*.

To each event corresponds a point of the space $\mathbb{R}^{1,3}$. The transition to another inertial reference system $\tilde{x}^0 = c\tilde{t}, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3$ is specified by an element of the *Poincaré group* (the group of motions of the space $\mathbb{R}^{1,3}$). This group is 10-dimensional and is generated by translations and "rotations" from $\mathrm{O}(1,3)$.

A vector $\xi$ is said to be time-like, light-like, or space-like if $\langle\xi,\xi\rangle > 0$, $\langle\xi,\xi\rangle = 0$, or $\langle\xi,\xi\rangle < 0$, respectively. A curve $\gamma(\tau)$ is called time-like, light-like, or space-like depending on the type of tangent vectors $u = \frac{d\gamma}{d\tau}$ to this curve.

To each point particle corresponds its world line $\gamma(\tau)$ in $\mathbb{R}^{1,3}$, which describes the states of this particle. For a massive particle, the world line is time-like. Particles without mass (e.g., photons) propagate along light-like curves.

To describe the motion of a free particle with mass $m > 0$, one of the following two functionals is used:

$$S_1 = \int_{\gamma(\tau)} L_1\, d\tau = \frac{m}{2}\int_{\gamma(\tau)} \langle u, u\rangle\, d\tau,$$
$$S_2 = \int_{\gamma(\tau)} L_2\, d\tau = -mc\int_{\gamma(\tau)} \sqrt{\langle u, u\rangle}\, d\tau,$$

where $u = \frac{d\gamma}{d\tau}$ is the velocity vector.

The functional $S_2$ is proportional to the length functional in $\mathbb{R}^{1,3}$ and hence is independent of the parametrization of the curve $\gamma(\tau)$. Therefore, we may parametrize $\gamma$ by the world time $\tau = t$ and rewrite $S_2$ as

$$S_2 = -mc^2\int\sqrt{1-\frac{v^2}{c^2}}\, dt,$$

where $u = (c, v^1, v^2, v^3)$, $v = \left(\frac{dx^1}{dt}, \frac{dx^2}{dt}, \frac{dx^3}{dt}\right)$ is the three-dimensional velocity vector, and $v^2 = \sum(v^\alpha)^2$.

The energy of the particle is equal to

$$E = v^\alpha\frac{\partial L_2}{\partial v^\alpha} - L_2 = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}},$$

and the three-dimensional momentum $\mathbf{p} = (p_1, p_2, p_3)$ has the form

$$p_\alpha = \frac{\partial L_2}{\partial v^\alpha} = \frac{mv^\alpha}{\sqrt{1-\frac{v^2}{c^2}}}.$$

For $v = 0$, we obtain the famous formula for the energy of the rest mass:

$$E = mc^2. \tag{15.1}$$

In the nonrelativistic limit as $|v| \ll c$, the Lagrangian $L_2$ becomes, up to a constant factor, the usual kinetic energy:

$$L_2 = -mc^2\sqrt{1-\frac{v^2}{c^2}} = -mc^2 + \frac{mv^2}{2} + O\Big(\frac{v^4}{c^2}\Big) \approx -mc^2 + \frac{mv^2}{2},$$

and the energy and momentum become

$$E \approx mc^2 + \frac{mv^2}{2}, \quad p_\alpha \approx mv^\alpha.$$

Thus in the nonrelativistic limit they take the classical form (with addition of a constant term $mc^2$ to the energy).

We see that free particles propagate along geodesics of the Minkowski metric, and the equations of motion in the nonrelativistic limit turn into the equations of classical mechanics.

In the four-dimensional formalism, where no world time $t$ is singled out, it is more expedient to use the functional $S_1$. Then all the notions can be extended to the case of a general metric $g_{ik}$ of signature $(+---)$ in general relativity theory. It follows from the Euler–Lagrange equations for the functional $S_1$ that the quantity $u^i \frac{\partial L_1}{\partial u^i} - L_1 = \frac{m}{2}\langle u, u\rangle$ is constant along the extremals, hence the parameter $\tau$ on them is natural, $d\tau = \text{const} \cdot dl$. An extremal of the functional $S_1$ taken relative to any other natural parameter is again an extremal, so that we set $d\tau = \frac{dl}{c}$, i.e.,

$$\left|\frac{dl}{d\tau}\right|^2 = \langle u, u\rangle = c^2, \quad u^i = \frac{dx^i}{d\tau}.$$

The parameter $\tau$ on the curve (world line of a particle) is called the particle's *proper time* if it satisfies this relation: $c = \frac{dl}{d\tau}$. The velocity vector $u$ relative to this parameter is called the (invariant) *velocity 4-vector*:

$$u = \left(\frac{c}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{v}{\sqrt{1-\frac{v^2}{c^2}}}\right), \quad v = (v^1, v^2, v^3).$$

The four-dimensional momentum (a covector, with subscripts) equals

$$p_* = \frac{\partial L_1}{\partial u} = \left(\frac{E}{c}, -\mathbf{p}\right),$$

and the momentum 4-vector is obtained from it by index raising in terms of the Minkowski metric:

$$p^i = g^{ik} p_k = \left(\frac{E}{c}, \mathbf{p}\right).$$

This vector satisfies the equation

$$g_{ij} p^i p^j = (p^0)^2 - \sum_{\alpha=1}^{3} (p^\alpha)^2 = \frac{E^2}{c^2} - \mathbf{p}^2 = m^2 c^2. \tag{15.2}$$

In the space of momenta $p$ with Minkowski metric $g_{ij}$ this equation specifies a mass surface, which is a three-dimensional Lobachevsky space (of constant negative curvature). We discussed a similar specification of a Lobachevsky plane in Section 4.3.

The energy as a function of momenta and coordinates is a Hamilton function. We derive from (15.2) that

$$H = c\sqrt{\mathbf{p}^2 + m^2c^2},$$

and in the nonrelativistic limit $|\mathbf{p}| \ll mc$ we again obtain the classical expression for the Hamiltonian (up to addition of a constant term):

$$H \approx mc^2 + \frac{\mathbf{p}^2}{2m}.$$

The inclusion of an electromagnetic field with potential $A_i(x)\,dx^i$ consists in introducing additional terms into the Lagrangians:

$$L_1 \to L_1 + \frac{e}{c}\,A_i\,\frac{dx^i}{d\tau}, \quad \tau = \frac{l}{c},$$
$$L_2 \to L_2 - \frac{e}{c}\,A_\alpha\,\frac{dx^\alpha}{dt} - eA_0, \quad t = \frac{x^0}{c},$$

where $e$ is the charge of the particle. This is equivalent to a translation of the momentum 4-vector (see Section 12.4),

$$p^i \to p^i + \frac{e}{c}A^i(x), \quad A^i = g^{ik}A_k, \quad i = 0, 1, 2, 3.$$

The energy of the particle in the electromagnetic field is equal to

$$E = v^\alpha\,\frac{\partial L_2}{\partial v^\alpha} - L_2 = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + eA_0,$$

and the energy as a function of coordinates and the three-dimensional momentum is

$$H(x,p) = c\sqrt{\sum_{\alpha=1}^{3}\Big(p_\alpha - \frac{e}{c}\,A_\alpha\Big)^2 + m^2c^2} + eA_0(x).$$

The action of the field itself is expressed in terms of the tensor $F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}$ as follows:

$$S(F) = -\frac{1}{16\pi c}\int F_{ik}F^{ik}\,d^4x.$$

Obviously,

$$d^2(A_ix^i) = d(F_{ik}\,dx^i \wedge dx^k) = \Big(\frac{\partial F_{ij}}{\partial x^k} + \frac{\partial F_{jk}}{\partial x^i} + \frac{\partial F_{ki}}{\partial x^j}\Big)\,dx^i \wedge dx^j \wedge dx^k = 0$$

(the first pair of Maxwell's equations). By varying the functionals $S(F)+S_1$ or $S(F)+S_2$ (obtained by adding the terms determined by the field $A$ to the Lagrangians) in components of the potential $A_i$, we obtain the Euler–Lagrange equations which form the second pair of the Maxwell equations:

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c}\,j^i,$$

where $j$ is the current vector vanishing in vacuum (see Section 14.1).

The equations of motion for multiparticle systems are the Euler–Lagrange equations for the functionals $S(F) + \sum_k S_1^{(k)}$ or $S(F) + \sum_k S_2^{(k)}$,

where summation is extended over the particles of the system, and $S_i^{(k)}$, $i = 1, 2$, is the action functional for the $k$th particle defined above.

**15.1.2. Gravitation field as a metric.** The basic hypothesis of Einstein's general relativity theory (GRT) is that the gravitational field is a metric $g_{ik}$ of signature $(+ - - -)$ in the 4-dimensional space-time $M^4$ with coordinates $(x^0, x^1, x^2, x^3)$. In general, this metric is not flat and may have a nonzero curvature. The length element $dl$ in this space is specified by the formula

$$dl^2 = g_{ik}\, dx^i\, dx^k.$$

The motion of free (point) particles of positive mass $m > 0$ goes along time-like geodesics of this metric. Since we do not single out the time coordinate, it is more convenient to regard geodesics as extremals of the functional

$$S_1 = \frac{m}{2} \int \langle u, u \rangle\, d\tau = \frac{m}{2} \int g_{ik} \frac{dx^i}{d\tau} \frac{dx^k}{d\tau}\, d\tau. \tag{15.3}$$

The integration here is performed along the world line of the particle. Particles of zero mass propagate along light-like geodesics of the metric $g_{ik}$.

Since the formally calculated "energy" $\frac{\partial L}{\partial u^i} u^i - L = \langle u, u \rangle$ is constant along the extremals of the functional $S_1$, the parameter $\tau$ is natural (proportional to $dl$). By setting $d\tau = \frac{dl}{c}$ we obtain the quantity, which, as in SRT, is called the proper time of the particle.

Inclusion of an electromagnetic field with vector-potential $A_i(x)$ is done in the same way as in SRT, i.e., by addition of a term determined by interaction between the charged particle (with charge $e$) and the field to the action functional:

$$\frac{m}{2} \int \langle u, u \rangle\, d\tau \to \frac{m}{2} \int \langle u, u \rangle\, d\tau + \frac{e}{c} \int A_i u^i\, d\tau.$$

In the definition of the action of the electromagnetic field the Minkowski metric is replaced by the metric $g_{ik}$:

$$S = -\frac{1}{16\pi c} \int F_{ik} F^{ik} \sqrt{-g}\, d^4x = -\frac{1}{16\pi c} \int F_{ik} F_{lm} g^{il} g^{mk} \sqrt{-g}\, d^4x, \tag{15.4}$$

where $g = \det(g_{ik})$ and

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}.$$

The Maxwell equations become

$$\begin{aligned} &d(F_{ik}\, dx^i \wedge dx^k) = 0 \quad \text{(the first pair)}, \\ &\nabla_k F^{ik} = -\frac{4\pi}{c} j^i \quad \text{(the second pair)}, \end{aligned} \tag{15.5}$$

where $j$ is the current vector vanishing when there are no charges.

We see that all the functionals and equations stated above have the same form as in SRT, only with flat Minkowski metric replaced by the metric $g_{ik}$.

But now we cannot confine ourselves to the action functional describing only particles, electromagnetic field, and their interaction. To describe the system we must take into account the action of the gravitational field itself.

From the physical point of view it appears natural to impose the following conditions on the action functional of the gravitational field.

1) The Euler–Lagrange equation must have the same form in all reference systems (systems of coordinates). Mathematically this means that the equation must be of tensor nature.

2) The Euler–Lagrange equation must contain at most second-order derivatives of the field variables, i.e., of coefficients of the metric $g_{ik}$.

3) The Lagrange function $L$ must depend only on the coefficients $g_{ik}$ and their first-order derivatives.

4) For "weak" gravitational fields, the motion equations of a slow particle must become the Newton equations.

Conditions 2) and 3) require that the Lagrangians and Euler–Lagrange equations include the variables subject to variation in the same way as the functionals considered before (e.g., the energy functional for curves or the action of an electromagnetic field). We explain the meaning of the fourth condition.

Let $x^0 = ct$ (as in SRT), and let $\tau = \frac{l}{c}$ be the proper time along a geodesic. By a "weak" gravitational field we will mean a metric $g_{ik}$ which can be expanded in a series in small parameter $c^{-1}$,

$$g_{ik} = \overset{(0)}{g_{ik}} + \frac{1}{c^2}\overset{(2)}{g_{ik}} + \frac{1}{c^3}\overset{(3)}{g_{ik}} + \cdots = \overset{(0)}{g_{ik}} + O\Big(\frac{1}{c^2}\Big),$$

with leading term $\overset{(0)}{g_{ik}}$ being the Minkowski metric. We will call a particle "slow" if its three-dimensional velocity is much less than the velocity of light, $|v| \ll c$, where $v = \left(\frac{dx^\alpha}{dt}\right)$, $\alpha = 1, 2, 3$. This will be expressed by assuming that $\frac{|v|}{c} = O\left(\frac{1}{c}\right)$ in subsequent formal expansions.

By definition,

$$d\tau = \frac{dl}{c} = \sqrt{\frac{1}{c^2}\, g_{ik}\, \frac{dx^i}{dt}\frac{dx^k}{dt}}\; dt,$$

which implies that

$$d\tau = \sqrt{1 + O\Big(\frac{1}{c^2}\Big)}\; dt = \Big[1 + O\Big(\frac{1}{c^2}\Big)\Big]\, dt. \tag{15.6}$$

The derivatives $\frac{\partial g_{ik}}{\partial x^\alpha}$ for $\alpha = 1, 2, 3$ are of order $O(c^{-2})$, since the spacial coordinates are assumed to be finite. We have $x^0 = ct$; therefore, the derivatives $\frac{\partial g_{ik}}{\partial x^0}$ are of order $O(c^{-3})$, and the second-order derivatives of the metric with respect to $x^0$ are $O(c^{-4})$.

Now we recall that geodesics (extremals of the functional (15.3)) satisfy the equation

$$\frac{d^2x^i}{d\tau^2} + \Gamma^i_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau} = 0$$

and the Christoffel symbols are

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\Big(\frac{\partial g_{jl}}{\partial x^k} + \frac{\partial g_{kl}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l}\Big). \tag{15.7}$$

By (15.6), the equation of geodesics can be written, up to terms of order $O(c^{-2})$, as

$$\ddot{x}^\alpha + \Gamma^\alpha_{jk}\dot{x}^j\dot{x}^k = 0,$$

where a dot means differentiation with respect to $t$ and $\alpha = 1, 2, 3$. The second term is of maximal order as $j = k = 0$:

$$\Gamma^\alpha_{jk}\dot{x}^j\dot{x}^k = \Gamma^\alpha_{00}c^2 + O\Big(\frac{1}{c}\Big).$$

Moreover, we derive from the general formula for the Christoffel symbols that

$$\Gamma^\alpha_{00} = -\frac{1}{2}g^{\alpha\alpha}\frac{\partial g_{00}}{\partial x^\alpha} + O\Big(\frac{1}{c^3}\Big) = \frac{1}{2}\frac{\partial g_{00}}{\partial x^\alpha} + O\Big(\frac{1}{c^3}\Big).$$

Then the equation of geodesics becomes

$$\ddot{x}^\alpha = -\Gamma^\alpha_{00}c^2 + O\Big(\frac{1}{c}\Big) = -\frac{1}{2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\Big(\frac{1}{c}\Big). \tag{15.8}$$

We recall the Newton equation describing the motion of a particle in a gravitational field:

$$\ddot{x}^\alpha = -\frac{\partial\varphi}{\partial x^\alpha},$$

where $\varphi$ is the potential of the gravitational field, which, outside the masses giving rise to the field, satisfies the Laplace equation

$$\Delta\varphi = \sum_\alpha \frac{\partial^2\varphi}{(\partial x^\alpha)^2} = 0.$$

Comparing the Newton equation with (15.8), we obtain the following assertion.

**Lemma 15.1.** *A necessary condition for the equations of motion of a slow particle in a weak gravitational field to coincide with the Newton equations within $O(c^{-1})$ is that the coefficient $\varphi$ in the expansion*

$$g_{00} = 1 + \frac{2\varphi(x)}{c^2} + O\Big(\frac{1}{c^3}\Big)$$

*satisfies the Laplace equation $\Delta\varphi = 0$, i.e., $\varphi$ is the potential of a weak gravitational field.*

**15.1.3. The action functional of a gravitational field.** Let $g_{ik}$ be a gravitational field in the 4-dimensional space $M^4$. We write down the corresponding Ricci tensor

$$R_{ik} = \frac{\partial \Gamma^l_{ik}}{\partial x^l} - \frac{\partial \Gamma^l_{il}}{\partial x^k} + \Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{km}$$

and the scalar curvature $R = g^{ik}R_{ik}$. In GRT the action of the gravitational field is taken to be

$$S_g = -\frac{c^3}{16\pi G}\int R\sqrt{-g}\,d^4x, \tag{15.9}$$

where $G = 6.67 \cdot 10^{-8}\,\mathrm{cm}^3/\mathrm{g}\cdot\mathrm{sec}^2$ is the gravitational constant.

As was shown in 14.2.3, the Euler–Lagrange equations for this functional are the *Einstein equations* (in vacuum)

$$R_{ik} - \frac{1}{2}Rg_{ik} = 0, \tag{15.10}$$

which are equivalent to $R_{ik} = 0$. These equations have a tensor structure, hence they look the same in any coordinate system. Note also that the expressions for the Christoffel symbols involve only the coefficients $g_{ik}$ and their first-order derivatives, hence equations (15.10) contain only the coefficients $g_{ik}$ with their first- and second-order derivatives.

We show that the functional (15.9) also satisfies the other two physical conditions stated in 15.1.2.

**Lemma 15.2.** *The following identity holds:*

$$R\sqrt{-g} = L\sqrt{-g} + \Big(\frac{\partial}{\partial x^l}(\sqrt{-g}g^{ik}\Gamma^l_{ik}) - \frac{\partial}{\partial x^k}(\sqrt{-g}g^{ik}\Gamma^l_{il})\Big),$$

*where*

$$L = g^{ik}(\Gamma^m_{il}\Gamma^l_{mk} - \Gamma^l_{ik}\Gamma^m_{lm}). \tag{15.11}$$

**Proof.** We have

$$R\sqrt{-g} = g^{ik}R_{ik}\sqrt{-g} = \sqrt{-g}\Big(g^{ik}\frac{\partial \Gamma^l_{ik}}{\partial x^l} - g^{ik}\frac{\partial \Gamma^l_{il}}{\partial x^k} + g^{ik}(\Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{km})\Big).$$

The first two sums can be transformed as

$$\sqrt{-g}g^{ik}\frac{\partial \Gamma^l_{ik}}{\partial x^l} = \frac{\partial}{\partial x^l}(\sqrt{-g}g^{ik}\Gamma^l_{ik}) - \Gamma^l_{ik}\frac{\partial}{\partial x^l}(\sqrt{-g}g^{ik}),$$
$$\sqrt{-g}g^{ik}\frac{\partial \Gamma^l_{il}}{\partial x^k} = \frac{\partial}{\partial x^k}(\sqrt{-g}g^{ik}\Gamma^l_{il}) - \Gamma^l_{il}\frac{\partial}{\partial x^k}(\sqrt{-g}g^{ik}).$$

Here we take out the first-order derivatives to obtain

$$R\sqrt{-g} = L\sqrt{-g} + \Big(\frac{\partial}{\partial x^l}(\sqrt{-g}g^{ik}\Gamma^l_{ik}) - \frac{\partial}{\partial x^k}(\sqrt{-g}g^{ik}\Gamma^l_{il})\Big),$$

where

$$
\begin{aligned}
L = {} & \Gamma^l_{il} \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^k} (\sqrt{-g} g^{ik}) \\
& - \Gamma^l_{ik} \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^l} (\sqrt{-g} g^{ik}) + g^{ik}(\Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{km}).
\end{aligned} \tag{15.12}
$$

We write down the first two terms as

$$I = \Gamma^l_{il} \frac{\partial g^{ik}}{\partial x^k} - \Gamma^l_{ik} \frac{\partial g^{ik}}{\partial x^l} + g^{ik}\Big(\Gamma^l_{il} \frac{1}{2g} \frac{\partial g}{\partial x^k} - \Gamma^l_{ik} \frac{1}{2g} \frac{\partial g}{\partial x^l}\Big).$$

The following relations are direct consequences of (15.7):

$$
\begin{aligned}
\Gamma^i_{ki} = g^{il} \frac{\partial g_{il}}{\partial x^k} &= \frac{1}{\sqrt{|g|}} \frac{\partial \log \sqrt{|g|}}{\partial x^k} = \frac{1}{2g} \frac{\partial g}{\partial x^k}, \\
\frac{\partial g^{ik}}{\partial x^l} &= -\Gamma^i_{ml} g^{mk} - \Gamma^k_{ml} g^{im}.
\end{aligned}
$$

After substituting them into the formula for $I$ and collecting the like terms, we obtain

$$I = 2g^{ik}(\Gamma^m_{il}\Gamma^l_{mk} - \Gamma^l_{ik}\Gamma^m_{lm}).$$

In turn, putting this expression into (15.12), we get $L = g^{ik}(\Gamma^m_{il}\Gamma^l_{mk} - \Gamma^l_{ik}\Gamma^m_{lm})$. Hence the lemma. □

Thus the action functional becomes

$$\int R\sqrt{-g}\, d^4x = \int L\sqrt{-g}\, d^4x + \int \frac{\partial}{\partial x^i} (\sqrt{-g} w^i)\, d^4x.$$

By the Stokes theorem the integral of the divergence $w$ reduces to the integral over a hypersurface enclosing the 4-dimensional domain of integration. If $w$ decreases rapidly at infinity (which holds, e.g., if the metric $g_{ik}$ converges to the Minkowski metric sufficiently fast), then the variation of the second term always vanishes, since we do not vary the metric on the "finite part" of the boundary of the domain. Hence, under these conditions, the Einstein equations are obtained as the Euler–Lagrange equations for the functional

$$\int L\sqrt{-g}\, d^4x,$$

where the Lagrangian $L$ depends only on the coefficients $g_{ik}$ and their first-order derivatives.

Since the Christoffel symbols do not form a tensor, $L$ is not a scalar, and it changes its value under changes of coordinates. An invariant quantity is provided by the scalar $R$ involved in the expression for the action of the gravitational field.

We now consider a weak gravitational field $g_{ik}$ satisfying the Einstein equation $R_{ik} = 0$. We will consider only the equation $R_{00} = 0$, which is written in detail as

$$R_{00} = \frac{\partial \Gamma^l_{00}}{\partial x^l} - \frac{\partial \Gamma^l_{0l}}{\partial x^0} + \Gamma^l_{00}\Gamma^m_{lm} - \Gamma^m_{0l}\Gamma^l_{0m} = 0.$$

Since $g_{ik} - g^{(0)}_{ik} = O(c^{-2})$, where $g^{(0)}_{ik}$ is the Minkowski metric, and $x^0 = ct$, we obtain

$$\frac{\partial g_{ik}}{\partial x^0} = O\Big(\frac{1}{c^3}\Big), \quad \frac{\partial g_{ik}}{\partial x^\alpha} = O\Big(\frac{1}{c^2}\Big), \quad \alpha = 1, 2, 3,$$

and consequently, $\Gamma^i_{jk} = O(c^{-2})$. Hence we conclude that

$$R_{00} = \frac{1}{2c^2} \sum_{\alpha=1}^{3} \frac{\partial^2 g^{(2)}_{00}}{(\partial x^\alpha)^2} + O\Big(\frac{1}{c^3}\Big).$$

Expanding $g_{00}$ in a series in powers of $c$ we get

$$g_{00} = 1 + \frac{2\varphi}{c^2} + O\Big(\frac{1}{c^3}\Big).$$

We substitute this expansion into the equation $R_{00} = 0$ to obtain

$$\frac{1}{c^2} \sum_{\alpha=1}^{3} \frac{\partial^2 \varphi}{(\partial x^\alpha)^2} + O\Big(\frac{1}{c^3}\Big) = 0,$$

which implies that $\varphi$ satisfies the Laplace equation $\Delta\varphi = 0$.

Therefore, the functional $\int R\sqrt{-g}\, d^4x$ satisfies all the conditions stated in 15.1.2.

The same conditions are fulfilled for the functional

$$\int (R + 2\Lambda)\sqrt{-g}\, d^4x,$$

where $\Lambda$ is the *cosmological constant*. The Euler–Lagrange equations for this functional are

$$R_{ik} - \frac{1}{2} R g_{ik} = \Lambda g_{ik}.$$

If we assume that the coordinates have physical dimension of length (are measured in centimeters), then the coefficients of the metric tensor $g_{ik}$ are dimensionless, and the scalars $R$ and $\Lambda$ are of physical dimension $1/\mathrm{cm}^2$. The physical experiments known by now estimate $\Lambda$ as $\Lambda \approx 10^{-62}\,\mathrm{cm}^{-2}$.

**15.1.4. The Schwarzschild and Kerr metrics.** In Newtonian mechanics the Poisson equation describes the potential $\varphi$ of the gravitational field produced by masses distributed in three-dimensional space with density $\rho$:

$$\Delta\varphi = 4\pi G\rho.$$

Outside the masses that produce the field, the Poisson equation becomes the Laplace equation $\Delta\varphi = 0$. The simplest spherically symmetric solution

$$\varphi = -G\frac{M}{r}, \quad r = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2},$$

describes the gravitational field produced by a point body of mass $M$ concentrated at the origin.

In GRT it is natural to seek an analog of this solution as a spherically symmetric metric $dl^2 = A\,dt^2 - B\,dr^2 - C(d\theta^2 + \sin^2\theta\,d\varphi^2)$, where $(r,\theta,\varphi)$ are spherical coordinates in $\mathbb{R}^3$. By an appropriate change of $r$ such a metric reduces to the form

$$dl^2 = e^\nu c^2\,dt^2 - e^\lambda\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\varphi^2).$$

We find the Christoffel symbols of this metric in coordinates $x^0 = ct$, $x^1 = r$, $x^2 = \theta$, $x^3 = \varphi$ by formulas (15.7):

$$\Gamma^1_{11} = \frac{\lambda'}{2}, \quad \Gamma^0_{10} = \frac{\nu'}{2}, \quad \Gamma^2_{33} = -\sin\theta\cos\theta, \quad \Gamma^0_{11} = \frac{\dot\lambda}{2}\,e^{\lambda-\nu},$$
$$\Gamma^1_{22} = -re^{-\lambda}, \quad \Gamma^1_{00} = \frac{\nu'}{2}\,e^{\nu-\lambda}, \quad \Gamma^2_{12} = \Gamma^3_{13} = \frac{1}{r}, \quad \Gamma^3_{23} = \cot\theta,$$
$$\Gamma^0_{00} = \frac{\dot\nu}{2}, \quad \Gamma^1_{10} = \frac{\dot\lambda}{2}, \quad \Gamma^1_{33} = -r\sin^2\theta e^{-\lambda},$$

where a dot means differentiation with respect to $ct$, and a prime means differentiation with respect to $r$. The Einstein equations $R_{ik} = 0$ become

$$\begin{aligned} &\dot\lambda = 0, \\ &e^{-\lambda}\Big(\frac{\nu'}{r} + \frac{1}{r^2}\Big) - \frac{1}{r^2} = 0, \\ &e^{-\lambda}\Big(\frac{\lambda'}{r} - \frac{1}{r^2}\Big) + \frac{1}{r^2} = 0. \end{aligned} \tag{15.13}$$

By summing up the last two equations, we obtain $\lambda + \nu = f(t)$. The change of time $\tilde t = \psi(t)$, where $\dot\psi = e^{f/2}$, transforms the coefficient $g_{00}$ of the metric by the formula

$$e^\nu c\,dt^2 = e^\nu(\dot\psi)^{-2}\,d\tilde t^2 = e^{\nu-f}\,d\tilde t^2,$$

while all the other coefficients of the metric remain unchanged. Hence we may assume without loss of generality that $\lambda + \nu = 0$. This implies that a spherically symmetric field in vacuum is always stationary (does not depend on time).

Equations (15.13) reduce to the equations

$$\dot\lambda = 0, \quad \lambda' = \frac{1 - e^\lambda}{r}.$$

The general solution to this first-order equation depends on a single constant $r_g$ and is given by the formula

$$e^{-\lambda} = 1 - \frac{r_g}{r}.$$

The corresponding metric is

$$dl^2 = \Big(1 - \frac{r_g}{r}\Big) c^2\, dt^2 - \frac{dr^2}{1 - \frac{r_g}{r}} - r^2 (d\theta^2 + \sin^2\theta\, d\varphi^2). \tag{15.14}$$

It is seen from the derivation of this formula that it provides the general form of a spherically symmetric metric satisfying the Einstein equations in vacuum. We set

$$r_g = \frac{2GM}{c^2}$$

and represent this metric as a weak gravitational field:

$$dl^2 = dl_0^2 - \frac{2GM}{c^2 r}\, d(ct)^2 - \frac{2GM}{c^2 r - 2GM}\, dr^2,$$

where $dl_0^2$ is the squared length element in the Minkowski metric. Taking into account the relationship between a weak gravitational field and the Newton potential (see Lemma 15.1), we conclude that, in cases of physical interest, $M > 0$ and the metric

$$dl^2 = \Big(1 - \frac{2GM}{c^2 r}\Big) c^2\, dt^2 - \frac{dr^2}{1 - \frac{2GM}{c^2 r}} - r^2 (d\theta^2 + \sin^2\theta\, d\varphi^2),$$

which is called the *Schwarzschild metric*, describes a gravitational field produced by a spherically symmetric distribution of masses. The mass of the body that produces the field is equal to $M$. The quantity $r_g$ is called the *Schwarzschild gravitational radius* (for the mass of the Earth, $r_g = 0.44$ cm, and for the mass of the Sun, $r_g = 3$ km).

If the density of a body is so high that the size of the body is less than or of the same order as the Schwarzschild radius $r_g$, then formula (15.14) for the gravitational field of this body has a singularity as $r \to r_g$. This singularity is removed by transition to new coordinates in which the metric depends on time. We will not discuss this issue here, and only point out that formula (15.14) is correct in the domain $r > r_g$.

Explicit formulas for the Schwarzschild metric enable us to demonstrate an important physical phenomenon, which was predicted by GRT, namely, the deviation of a light beam due to a gravitational field. Recall that light propagates along light-like geodesics specified by the equations

$$\frac{d^2 x^i}{d\tau^2} + \Gamma^i_{jk} \frac{dx^j}{d\tau} \frac{dx^k}{d\tau} = 0, \quad i, j, k = 0, 1, 2, 3, \quad g_{ik} \frac{dx^i}{d\tau} \frac{dx^k}{d\tau} \equiv 0.$$

The general form of solution to these equations is

$$\varphi = \int \frac{dr}{r^2\sqrt{\frac{1}{\rho^2} - \frac{1}{r^2}\left(1 - \frac{r_g}{r}\right)}} + \varphi_0, \qquad \rho = \text{const},$$

where $(r, \varphi)$ are polar coordinates in the plane in which the geodesic lies. As $r_g \to 0$, this solution turns into the straight line $r\cos(\varphi - \varphi_0) = \rho$, passing at the distance $\rho$ from the center ($r = 0$) and going to infinity at the angles $\varphi_0 \pm \pi/2$.

Let $f(\tau) = (r(\tau), \theta(\tau))$ be a light-like geodesic and $\rho \neq 0$. Asymptotically, as $\tau \to \pm\infty$, it behaves as the straight line $r\cos(\varphi - \varphi_\pm) = \rho$. The geodesic is curvilinear and bent towards the center ("is attracted by the center"). The exact formula for the absolute value of the difference $|\varphi_+ - \varphi_-|$ is rather complicated, but by a series expansion in small parameter $r_g/\rho$ we obtain that $|\varphi_+ - \varphi_-|$ differs from $\pi$ by

$$\delta\varphi \approx \frac{2r_g}{\rho}.$$

The Schwarzschild metric specifies a spherically symmetric solution to the Einstein equations in vacuum, which asymptotically (as $r \to \infty$) behaves as the Minkowski metric, i.e., is asymptotically flat. From the viewpoint of modern physics, this solution describes a stationary black hole.

Schwarzschild's solution is contained in a one-parameter family of axially symmetric asymptotically flat solutions to the Einstein equation in vacuum. They are called *Kerr's metrics* and describe uniformly rotating (around the $Oz$ axis) black holes. The explicit formula for such metrics is as follows:

$$dl^2 = c^2\,dt^2 - \frac{r_g r}{\rho^2}\,(a\sin^2\theta\,d\varphi + c\,dt)^2 - \frac{\rho^2}{\Delta}\,dr^2 - (r^2 + a^2)\sin^2\theta\,d\varphi^2 - \rho^2\,d\theta^2,$$

where $\Delta = r^2 - r_g r + a^2$, $\rho^2 = r^2 + a^2\cos^2\theta$. The quantity $a$ is proportional to the angular momentum, and for $a = 0$, Kerr's metric turns into Schwarzschild's.

Unlike the formula for the Schwarzschild metric, this formula may be correct for all positive $r$. This is true for large angular momenta when $r^2 - 2r_g r + a^2 > 0$, i.e., for $a^2 > \frac{r_g^2}{4}$.

**15.1.5. Interaction of matter with gravitational field.** Interaction of the gravitational field with all other fields and particles is described by the Euler–Lagrange equations for the total action functional $S = S_g + S_m$:

$$\frac{\delta S_g}{\delta g^{ik}} = -\frac{\delta S_m}{\delta g^{ik}}. \tag{15.15}$$

Here

$$S_g = -\frac{c^3}{16\pi G}\int R\sqrt{-g}\,d^4x$$

is the action functional for the gravitational field, and $S_m$ is the action functional for the other fields and particles (action functional of matter). For example, for the electromagnetic field we have

$$S_m = -\frac{1}{16\pi c}\int F_{ik}F^{ik}\sqrt{-g}\,d^4x.$$

In the general relativity theory it is assumed that the interaction of the gravitational field with matter is expressed by the energy-momentum tensor $T_{ik}$ of the matter, which is proportional to the right-hand side of (15.15):

$$R_{ik} - \frac{1}{2}Rg_{ik} = \mu T_{ik}, \quad \mu = \text{const}, \tag{15.16}$$

where $\mu$ is a universal constant.

The total system of the Euler–Lagrange equations has the form

$$\frac{\delta S}{\delta g^{ik}} = 0, \quad \frac{\delta S}{\delta \psi_\alpha} = \frac{\delta S_m}{\delta \psi_\alpha} = 0,$$

where the $\psi_\alpha$ are field variables describing the matter. For example, in the case of electromagnetic field, these are the coefficients of the vector-potential $A_i$, $i = 0, 1, 2, 3$.

Now we will find the value of the constant $\mu$ in (15.16). To this end, consider a dust cloud in a weak gravitational field. Assuming that the substance in the cloud has zero pressure and velocity, we find its energy-momentum tensor as

$$T_{ik} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & 0 & \\ 0 & & & \end{pmatrix},$$

where $\rho$ is the mass density. As we pointed out before (see 14.1.2), $T_{00}$ plays the role of the energy density. Therefore, the formula $T_{00} = \rho c^2$ is a direct consequence of the relativistic formula (15.1) for the energy of a mass at rest, $E = mc^2$.

We know that $g_{00} = 1 + 2\varphi/c^2 + O(1/c^3)$, where $\varphi$ is the gravitational potential satisfying the Poisson equation

$$\Delta\varphi = 4\pi G\rho$$

(see 15.1.3). Taking the trace in the left- and right-hand sides of (15.16), we obtain $R = R^i_i = -\mu T = -\mu T^i_i$. Using this, rewrite (15.16) as

$$R_{ik} = \mu\Big(T_{ik} - \frac{1}{2}Tg_{ik}\Big).$$

In particular,

$$R_{00} = \mu\Big(T_{00} - \frac{1}{2}\,Tg_{00}\Big).$$

For a weak gravitational field,

$$R_{00} = \frac{\Delta\varphi}{c^2} + O\Big(\frac{1}{c^3}\Big)$$

and $T = \rho c^2(1 + O(c^{-2}))$. Hence we obtain

$$\frac{\Delta\varphi}{c^2} + O\Big(\frac{1}{c^3}\Big) = \mu\,\frac{\rho c^2}{2}\Big(1 + O\Big(\frac{1}{c^2}\Big)\Big).$$

Since this equation must turn into the Poisson equation as $c \to \infty$, we conclude that the universal constant equals

$$\mu = \frac{8\pi G}{c^4}.$$

This formula and equations (15.15) and (15.16) imply the formula for the energy-momentum tensor in GRT. Indeed, since

$$\frac{\delta S_g}{\delta g^{ik}} = -\frac{c^3}{16\pi G}\sqrt{-g}\Big(R_{ik} - \frac{1}{2}\,Rg_{ik}\Big) = -\frac{1}{2c}\sqrt{-g}T_{ik} = -\frac{\delta S_m}{\delta g^{ik}},$$

we have

$$T_{ik} = \frac{2c}{\sqrt{-g}}\,\frac{\delta S_m}{\delta g^{ik}}.$$

This definition of the energy-momentum tensor immediately provides a symmetric tensor $T_{ik}$.

The energy-momentum tensor of electromagnetic field defined in this way coincides with the one obtained before in 14.2.2:

$$T_{ik} = \frac{1}{4\pi}\Big(-F_{il}F^l_k + \frac{1}{4}\,g_{ik}F_{lm}F^{lm}\Big).$$

The energy-momentum tensor of isotropic dense medium (the *hydrodynamic energy-momentum tensor*) has the form

$$T_{ik} = (p+\varepsilon)u_iu_k - pg_{ik},$$

where $p$ and $\varepsilon$ are pressure and energy density at each point of the medium in the "accompanying" coordinate system, relative to which this point stays at rest (i.e., the 4-vector of velocity $u$ has the form $u_0 = 1, u_\alpha = 0, \alpha = 1, 2, 3$). In the "accompanying" coordinate system, this tensor at the given point becomes

$$T_{ik} = \begin{pmatrix} \varepsilon & & & 0 \\ & p & & \\ & & p & \\ 0 & & & p \end{pmatrix}. \tag{15.17}$$

The Einstein equations are completed by the state equations establishing the relationship between $p$ and $\varepsilon$:

a) for dust substance $p = 0$, $\varepsilon = \rho c^2$;

b) the "ultra-relativistic" state equation is $p = \varepsilon/3$ or $T = T^a_a \equiv 0$.

Equations (15.16) and the Bianchi identity imply the identity

$$\nabla_k T^k_i = 0, \tag{15.18}$$

which replaces the conservation laws

$$\frac{\partial T^k_i}{\partial x^k} = 0$$

in a curvilinear coordinate system. However, the identity (15.18) does not correspond to any conservation law because in general it is not a condition of vanishing for the exterior derivative of a 3-form, and so no analog of Theorem 14.3 holds for this identity.

**15.1.6. On the concept of mass in general relativity theory.** The definition of the energy-momentum tensor of matter as $T_{ik} = \frac{\delta S_m}{\delta g^{ik}}$ does not work for a system involving both matter and the gravitational field. This follows from the fact that the equation of motion has the form $\frac{\delta S}{\delta g^{ik}} = 0$, i.e., the energy-momentum tensor so defined is identically equal to zero.

One could formally derive the energy-momentum tensor of the gravitational field in vacuum using the procedure presented in 14.1.2 and applicable to Lagrangians depending on field variables and their first-order derivatives. To this end one must replace the Lagrangian $R\sqrt{-g}$ by $L\sqrt{-g}$, where $L$ has the form (15.11). But in this case the Einstein equations also imply that the tensor $T_{ik}$ is identically equal to zero.

This difficulty is of fundamental nature, and, as a result of an extensive research, the following conclusions were made:

1) For a gravitational field that interacts with matter, the familiar concepts, like energy and momentum, are well defined if the metric tends to the flat Minkowski metric (the space is asymptotically flat) at infinity.

2) There is no uniquely defined energy density even for asymptotically flat spaces: energy cannot be localized.

Now we will describe how energy and momentum are defined for asymptotically flat spaces.

Let $M^4$ be the four-dimensional space-time with coordinates $x^0$, $x^1$, $x^2$, $x^3$ and metric $g_{ik}$ of signature $(+---)$. Assume that the surfaces $x^0 = \text{const}$ are space-like and $g_{00} > 0$ (the coordinate $x^0$ corresponds to time). Moreover, we assume that the metric $g_{ik}$ is asymptotically flat at

infinity, namely,

$$g_{ik} = g_{ik}^{(0)} + h_{ik},$$

where $g_{ik}^{(0)}$ is a Minkowski metric and

$$h_{ik} = O\Big(\frac{1}{r}\Big), \quad \frac{\partial h_{ik}}{\partial x^l} = O\Big(\frac{1}{r^2}\Big), \quad \frac{\partial^2 h_{ik}}{\partial x^l \partial x^m} = O\Big(\frac{1}{r^3}\Big),$$

$r = \sqrt{\sum_{\alpha=1}^{3}(x^\alpha)^2}$.

Assume that the metric satisfies the Einstein equations, which can be rewritten as

$$T^{ik} = \frac{c^4}{8\pi G}\Big(R^{ik} - \frac{1}{2}\,Rg^{ik}\Big).$$

Define the quantity

$$\eta^{ikl} = \frac{c^4}{16\pi G}\,\frac{\partial}{\partial x^m}\,[(-g)(g^{ik}g^{lm} - g^{il}g^{km})],$$

which is skew-symmetric in superscripts $k, l$.

Fix a point in the space and take geodesic coordinates in its neighborhood. The first-order derivatives of $g_{ik}$ at this point vanish, hence the Ricci tensor becomes

$$R^{ik} = \frac{1}{2}\,g^{im}g^{kp}g^{ln}\Big(\frac{\partial^2 g_{lp}}{\partial x^m \partial x^n} + \frac{\partial^2 g_{mn}}{\partial x^l \partial x^p} - \frac{\partial^2 g_{ln}}{\partial x^m \partial x^p} - \frac{\partial^2 g_{mp}}{\partial x^l \partial x^n}\Big),$$

and the energy-momentum tensor of matter takes the form

$$T^{ik} = -\frac{1}{g}\,\frac{\partial \eta^{ikl}}{\partial x^l}.$$

Since $\eta^{ikl} = -\eta^{ilk}$, at this point in these coordinates the following equality holds:

$$\frac{\partial(-g)T^{ik}}{\partial x^k} = 0, \tag{15.19}$$

which is a conservation law.

In general the difference

$$\mathbf{t}^{ik} = -\frac{1}{g}\Big(\frac{\partial \eta^{ikl}}{\partial x^l} - (-g)T^{ik}\Big)$$

is not equal to zero. It is called the *energy-momentum pseudotensor* and transforms as a tensor only under affine changes of coordinates. The $\mathbf{t}^{ik}$ can be expressed explicitly as homogeneous polynomials in $g^{ik}$ and Christoffel symbols $\Gamma^i_{jk}$, but we will not state this formula.

The definition of $\mathbf{t}^{ik}$ implies the identity

$$\frac{\partial}{\partial x^k}\,(-g(T^{ik} + \mathbf{t}^{ik})) = 0.$$

Define the momentum 4-vector by the formula

$$P^i = \frac{1}{c} \int_{x^0=\text{const}} (-g)(T^{ik} + \mathbf{t}^{ik})\, dS_k, \tag{15.20}$$

where $dS_k = (-1)^k\, dx^0 \wedge \cdots \wedge \widehat{dx^k} \wedge \cdots \wedge dx^3$. When there is no gravitational field, this vector becomes the 4-vector of material momentum defined in 14.1.2. Since the integrand is a total differential,

$$\frac{\partial \eta^{ikl}}{\partial x^l}\, dS_k = \frac{1}{2}\, \eta^{ikl}\, df_{kl},$$

by the Stokes theorem the integral (15.20) reduces to the boundary integrals over the spheres $S_R$ of radius $R$:

$$P^i = \frac{1}{2c} \lim_{R\to\infty} \int_{S_R} \eta^{ikl}\, df_{kl}. \tag{15.21}$$

For the momentum 4-vector $P$ the following analog of Theorem 14.3 holds.

**Theorem 15.1.** *Identity* (15.19) *implies that the momentum vector $P$ is preserved in time: $P(x_1^0) = P(x_2^0)$.*

When defining the momentum 4-vector we used special coordinates in which the metric $g_{ik}$ was asymptotically flat. If we consider a change of coordinates $x \to \tilde{x}$, then the requirement that in the new coordinate system the metric $\tilde{g}_{ik}$ be asymptotically flat imposes no restriction on the transformation $x \to \tilde{x}$ in a finite domain, but for large $r$ the following conditions must be fulfilled:

$$\tilde{x}^i = a^i_k x^k + b^i + O\Big(\frac{1}{r}\Big), \quad \frac{\partial \tilde{x}^i}{\partial x^k} = O\Big(\frac{1}{r^2}\Big), \quad \frac{\partial^2 \tilde{x}^i}{\partial x^k \partial x^l} = O\Big(\frac{1}{r^3}\Big), \tag{15.22}$$

i.e., the transformation of coordinates must behave asymptotically as a transformation $x \to Ax + b$, $A \in \mathrm{O}(1,3)$, from the Poincaré group. The following statement is now obvious.

**Lemma 15.3.** *If a change of coordinates $x \to \tilde{x}$ converges asymptotically to a transformation $x \to Ax + b$ from the Poincaré group, namely, if it satisfies conditions* (15.22), *then the momentum 4-vector transforms as a vector: $P \to AP$.*

Note that one can define also the angular momentum $M^{ik}$, which behaves as a tensor relative to the transformations of coordinates satisfying conditions (15.22).

Now we impose the following condition on the energy-momentum tensor $T_{ik}$.

*For any time-like vector $u$ at any space point, the vector $T^k_i u^i$ is not space-like and $T_{ik} u^i u^k \geq 0$.*

For example, in the case of the energy-momentum tensor (15.17) of a dust cloud in the Minkowski space, this condition means that the density of matter $\varepsilon$ is nonnegative at each point.

The following theorem is referred to as the *positive-mass theorem.*

**Theorem 15.2.** *Suppose that an asymptotically flat metric $g_{ik}$ and an energy-momentum tensor $T_{ik}$ satisfy the Einstein equations and the tensor $T_{ik}$ satisfies the condition stated above. Then the following inequality holds:*

$$(P^0)^2 - \sum_{\alpha=1}^{3}(P^\alpha)^2 \geq 0.$$

*Moreover, if $P^0 = 0$ for a hypersurface $x^0 = \text{const}$, then the induced metric $-g_{\alpha\beta}$ on this hypersurface is Euclidean.*

Recall that the self-energy of a relativistic particle in the Minkowski space equals $mc^2 = c\sqrt{(P^0)^2 - \sum_{\alpha=1}^{3}(P^\alpha)^2}$. Therefore, the positive-mass theorem shows that the energy of a system consisting of interacting gravitational field and matter is well defined in cases of physical interest (the "mass" $m$ is positive). Equality (15.21) implies that for nontrivial gravitational fields the metric $g_{ik}$ cannot converge to the Minkowski metric too fast: if

$$g_{ik} - g_{ik}^{(0)} = O\left(\frac{1}{r^2}\right),$$

then the restrictions of the metric $-g_{ik}$ to the surfaces $x^0 = \text{const}$ are Euclidean.

## 15.2. Spinors and the Dirac equation

**15.2.1. Automorphisms of matrix algebras.** As usual, let $\mathrm{M}(n, \mathbb{C})$ be the algebra of all complex $n \times n$ matrices. The mapping

$$h\colon \mathrm{M}(n, \mathbb{C}) \to \mathrm{M}(n, \mathbb{C})$$

is called an *automorphism* if it is an isomorphism of the algebra onto itself, i.e., it preserves all the operations:

$$h(AB) = h(A)h(B), \quad h(A+B) = h(A) + h(B), \quad h(\lambda A) = \lambda h(A)$$

for any $A, B \in \mathrm{M}(n, \mathbb{C})$, $\lambda \in \mathbb{C}$, and is invertible. For example, any invertible matrix $X \in \mathrm{M}(n, \mathbb{C})$ specifies an *inner automorphism*

$$A \to XAX^{-1}.$$

It turns out that the matrix algebras $\mathrm{M}(n, \mathbb{C})$ have no other automorphisms.

**Theorem 15.3.** *All the automorphisms of the matrix algebras $\mathrm{M}(n, \mathbb{C})$ are inner.*

**Proof.** A matrix $P \in \mathrm{M}(n, \mathbb{C})$ is called a *projector* if $P^2 = P$. Two projectors $P$, $Q$ are said to be *orthogonal* if $PQ = QP = 0$.

Let $e_1, \dots, e_n$ be a basis in $\mathbb{C}^n$. Define orthogonal projectors $P_1, \dots, P_n$ by their action on an arbitrary vector:

$$P_i(\xi^1 e_1 + \cdots + \xi^n e_n) = \xi^i e_i$$

(here in the right-hand side no summation over $i$ is assumed). We have

$$P_i^2 = P_i, \quad P_i P_j = 0 \quad \text{for } i \neq j, \quad P_1 + \cdots + P_n = 1. \tag{15.23}$$

Let $h \colon \mathrm{M}(n, \mathbb{C}) \to \mathrm{M}(n, \mathbb{C})$ be an automorphism. Set $h(P_i) = P_i'$. Since $h(1) = 1$, relations (15.23) become

$$(P_i')^2 = P_i', \quad P_i' P_j' = 0 \quad \text{for } i \neq j, \quad P_1' + \cdots + P_n' = 1.$$

Therefore, the $P_i'$ are pairwise orthogonal projectors. Since $P_1' + \cdots + P_n' = 1$, all these projectors are one-dimensional.

Define the matrices $t_{ij} \in \mathrm{M}(n, \mathbb{C})$ by the formula

$$t_{ij}(e_k) = e_i \delta_{jk} = \begin{cases} e_i & \text{for } j = k, \\ 0 & \text{for } j \neq k. \end{cases}$$

Obviously, the following relations hold:

$$P_i = t_{ii}, \quad t_{ij} t_{kl} = \begin{cases} t_{il} & \text{for } j = k, \\ 0 & \text{for } j \neq k. \end{cases}$$

The automorphism $h$ transforms these matrices into $t_{ij}' = h(t_{ij})$.

Now we construct the basis $e_1', \dots, e_n'$ in $\mathbb{C}^n$ in the following way: take a nonzero (generating) vector $e_1'$ in $P_1'(\mathbb{C}^n)$ and set $e_i' = t_{i1}'(e_1')$.

Let $X$ be the transformation of the space $\mathbb{C}^n$ defined by its action on the basis:

$$X(e_i) = e_i'.$$

We will show that for any matrix $A \in \mathrm{M}(n, \mathbb{C})$,

$$h(A) = XAX^{-1}. \tag{15.24}$$

Every matrix $A$ uniquely decomposes in a sum

$$A = \sum_{i,j=1,\dots,n} a_{ij} t_{ij},$$

where $a_{ij}$ are the matrix entries. Hence it suffices to prove (15.24) for the matrices $t_{ij}$. We have

$$X t_{ij} X^{-1}(e_k') = X t_{ij}(e_k) = X(e_i \delta_{jk}) = \delta_{jk} e_i'.$$

Obviously,

$$h(t_{ij})(e_k') = t_{ij}'(e_k') = e_i' \delta_{jk}.$$

Thus we have shown that

$$h(t_{ij}) = X t_{ij} X^{-1}$$

for any $i, j = 1, \dots, n$. By linearity this implies that $h(A) = XAX^{-1}$ for all matrices $A$. Hence the theorem. □

**15.2.2. Spinor representation of the group** SO(3). The spinor representation of the group SO(3) uses a special representation of the algebra $\mathrm{M}(2, \mathbb{C})$ in terms of the *Pauli matrices* $\sigma_1, \sigma_2, \sigma_3$:

(15.25)

$$\sigma_x = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We have the following obvious lemma.

**Lemma 15.4.** 1. *The Pauli matrices $\sigma_1, \sigma_2, \sigma_3$ together with the identity matrix* 1 *form an additive basis in the algebra* $\mathrm{M}(2, \mathbb{C})$.

2. *There are the following relationships between the Pauli matrices:*

$$[\sigma_k, \sigma_l] = \sigma_k \sigma_l - \sigma_l \sigma_k = 2i\varepsilon_{klm}\sigma_m, \tag{15.26}$$

$$\{\sigma_k, \sigma_l\} = \sigma_k \sigma_l + \sigma_l \sigma_k = 2\delta_{kl}. \tag{15.27}$$

Recall (see 7.2.2) that

$$\varepsilon^k_{lm} = \varepsilon_{klm} = \begin{cases} 1 & \text{if the permutation } (k, l, m) \text{ is even,} \\ -1 & \text{if the permutation } (k, l, m) \text{ is odd,} \\ 0 & \text{if a pair of indices among } k, l, m \text{ coincide.} \end{cases}$$

Relations (15.26) imply that the matrices $\frac{i}{2}\sigma_j$ realize a representation of the Lie algebra of the group SO(3) (or SU(2)) in $\mathrm{M}(2, \mathbb{C})$.

Let $A\colon \mathbb{R}^3 \to \mathbb{R}^3$, $A \in \mathrm{SO}(3)$, be an orthogonal transformation:

$$(x^i) \to (x'^i), \quad x'^i = a^i_j x^j, \quad i = 1, 2, 3.$$

It induces a linear transformation of $\mathrm{M}(2, \mathbb{C})$ by the formula

$$A(\sigma_i) = \sigma'_i = a^j_i \sigma_j, \quad A(1) = 1. \tag{15.28}$$

**Theorem 15.4.** *The matrices $\sigma'_i$, $i = 1, 2, 3$, satisfy the relations* (15.26) *and* (15.27).

*Thus the mapping* (15.28) *specifies an automorphism*

$$h_A\colon \mathrm{M}(2, \mathbb{C}) \to \mathrm{M}(2, \mathbb{C})$$

*of the matrix algebra* $\mathrm{M}(2, \mathbb{C})$.

**Proof.** We have

$$\{\sigma'_k, \sigma'_l\} = a^p_k a^q_l \{\sigma_p, \sigma_q\} = a^p_k a^q_l 2\delta_{pq} = 2\sum_p a^p_k a^p_l = 2(a^\top)^k_p a^p_l = 2\delta_{kl},$$

since the matrix $A$ is orthogonal ($A^\top A = 1$).

Now we verify the second relation:

$$[\sigma'_k, \sigma'_l] = a^p_k a^q_l [\sigma_p, \sigma_q] = a^p_k a^q_l (2i\varepsilon^r_{pq}\sigma_r) = 2ia^r_m \varepsilon^m_{kl}\sigma_r = 2i\varepsilon^m_{kl}\sigma'_m.$$

For the proof of the last equality we have used the fact that orthogonal transformations in SO(3) preserve also the vector product. Hence for arbitrary vectors $\xi, \eta$ we have

$$a^p_k a^q_l \varepsilon^r_{pq} \xi^k \eta^l = \varepsilon^r_{pq}(A\xi)^p(A\eta)^q = [A\xi \times A\eta]^r = (A[\xi \times \eta])^r = a^r_m \varepsilon^m_{kl}\xi^k\eta^l.$$

Relations (15.26) and (15.27) imply that the products $\sigma_i\sigma_j$ are linearly expressible in terms of the Pauli matrices and the identity matrix. Therefore, any linear mapping of $\mathrm{M}(2, \mathbb{C})$ into itself preserving these relations specifies an automorphism of the matrix algebra. Hence the theorem. □

**Corollary 15.1.** *By Theorem* 15.3 *there exists a transformation* $g = g(A)\colon \mathbb{C}^2 \to \mathbb{C}^2$ *such that* $h_A(X) = gXg^{-1}$ *for any matrix* $X$.

The mapping

$$A \to g(A)$$

is called the *spinor representation* of the group SO(3) in the group $\mathrm{GL}(2, \mathbb{C})$.

This representation is multivalued: the matrix $g(A)$ is defined up to a nonzero factor $\lambda \in \mathbb{C}$.

If we require that $g(A) \in \mathrm{SL}(2, \mathbb{C})$, then the representation becomes two-valued, with the two values differing by factor $-1$. We have actually described this representation in 5.3.2: the group of unit quaternions $G = \mathrm{SU}(2)$ acts on imaginary quaternions by the formula

$$x \to qx\bar{q},$$

which realizes an exact representation of the group $\mathrm{SO}(3) = \mathrm{SU}(2)/\{\pm 1\}$, and $g(A) \in \mathrm{SU}(2)$, $A \in \mathrm{SO}(3)$.

The following statement holds.

**Lemma 15.5.** *Under the two-valued spinor representation* $g\colon \mathrm{SO}(3) \to \mathrm{SL}(2, \mathbb{C})$, *the rotation through an angle* $\theta$ *about the axis with direction vector* $n = (n_x, n_y, n_z)$, $n_x^2 + n_y^2 + n_z^2 = 1$, *goes into the transformation*

$$g(\theta, n) = \exp\Big\{-i\,\frac{\theta}{2}\,(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z)\Big\}.$$

We leave the proof of this lemma as a problem for the reader.

**15.2.3. Spinor representation of the group** $\mathrm{O}(1,3)$. Let $g_{ab} = \mathrm{diag}(1,-1,-1,-1)$ be the Minkowski metric. In the algebra $\mathrm{M}(4,\mathbb{C})$ select the elements $\gamma^0$, $\gamma^1$, $\gamma^2$, $\gamma^3$ satisfying the relations

$$\gamma^a\gamma^b + \gamma^b\gamma^a = 2g^{ab}\cdot 1. \tag{15.29}$$

This can be done by taking the block $4\times 4$ matrices involving the Pauli matrices (15.25),

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \gamma^k = \begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix}, \quad k = 1,2,3. \tag{15.30}$$

The following statement holds.

**Lemma 15.6.** *The matrices* $1$, $\gamma^a$, $\gamma^a\gamma^b$ $(a<b)$, $\gamma^a\gamma^b\gamma^c$ $(a<b<c)$, *and* $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$ *are linearly independent.*

*Hence the algebra over the field* $\mathbb{C}$ *with generators* $\gamma^0$, $\gamma^1$, $\gamma^2$, $\gamma^3$ *and relations* (15.29) *is isomorphic to the matrix algebra* $\mathrm{M}(4,\mathbb{C})$.

**Proof.** It follows from relations (15.29) that any product of elements $\gamma^a$ reduces to a linear combination of the elements stated in the lemma. The number of these elements equals the dimension dim $= 16$ of the algebra $\mathrm{M}(4,\mathbb{C})$. It may be shown by direct calculations that all these elements are linearly independent. Hence the lemma. □

Consider the Minkowski space $\mathbb{R}^{1,3}$ with coordinates $(x^0,x^1,x^2,x^3)$, and let $\Lambda = (\lambda^a_b)$ be a matrix, $\Lambda \in \mathrm{O}(1,3)$. Define the matrices

$$\gamma'^a = \lambda^a_b\gamma^b.$$

Since $\Lambda$ preserves the Minkowski metric, relations (15.29) imply that the matrices $\gamma'^a$ also satisfy these relations:

$$\{\gamma'^a,\gamma'^b\} = 2g^{ab}\cdot 1.$$

Therefore, the mapping

$$1 \to 1, \quad \gamma^a \to \gamma'^a$$

of the generators of the algebra $\mathrm{M}(4,\mathbb{C})$ specifies by Lemma 15.6 an automorphism

$$h_\Lambda\colon \mathrm{M}(4,\mathbb{C}) \to \mathrm{M}(4,\mathbb{C}).$$

Theorem 15.3 implies that this automorphism is inner:

$$h_\Lambda(X) = gXg^{-1}, \quad g = g(\Lambda).$$

The mapping

$$\Lambda \to g(\Lambda)$$

is called the *spinor representation* of the group $\mathrm{O}(1,3)$ in the group $\mathrm{GL}(4,\mathbb{C})$. It is multivalued and is defined up to multiplication by a nonzero complex number $\lambda$.

If we require that $g(\Lambda) \in \mathrm{SL}(4, \mathbb{C})$, then the representation becomes four-valued, being defined up to multiplication by $\pm 1$ and $\pm i$, which are the 4th roots of one.

Here we have to distinguish between the spinor representation of the group $\mathrm{O}(1,3)$, which has four connected components (see 2.2.2), and the spinor representations of its subgroups, the group $\mathrm{SO}(1,3)$ of orientation-preserving transformations and the group $\mathrm{SO}^+(1,3) \subset \mathrm{SO}(1,3)$ of transformations that preserve both the direction of time and orientation.

The group $\mathrm{SO}^+(1,3)$ is the connected component of the group $\mathrm{O}(1,3)$ that contains the identity element of the group. The spinor representation

$$\Lambda \to g(\Lambda) \in \mathrm{SL}(4, \mathbb{C})$$

takes it into two connected components, one of which contains the identity element of the group $\mathrm{SL}(4, \mathbb{C})$, and the mapping into this component is two-valued and defined up to multiplication by $-1$. Extension of the spinor representation to the other connected components of the group $\mathrm{O}(1,3)$ is done differently in different problems, depending on physical considerations.

The space $\mathbb{C}^4$, on which the spinor representation acts, is called the space of (4-component) *spinors*. We write them as 4-component column vectors.

Now we describe how the spinor representation acts on some important subgroups of $\mathrm{O}(1,3)$.

**Lemma 15.7.** 1. *The spinor representation maps the rotation through an angle $\theta$ about a unit vector $n = (n_1, n_2, n_3)$ in $\mathbb{R}^3$ (an element of $\mathrm{SO}(3) \subset \mathrm{O}(1,3)$) into*

$$g(\theta, n) = \exp\Big\{-i\,\frac{\theta}{2}\,(n_1\Sigma_1 + n_2\Sigma_2 + n_3\Sigma_3)\Big\},$$

*where* $\Sigma_j = \left(\begin{smallmatrix} \sigma_j & 0 \\ 0 & \sigma_j \end{smallmatrix}\right)$.

2. *The hyperbolic rotation through an imaginary angle $i\theta$ in the plane $(x^0, n)$, where $n$ is a three-dimensional unit vector (an elementary Lorentz transformation), under spinor representation becomes*

$$g(\theta, n) = \exp\Big\{-\frac{\theta}{2}\,(n_1\alpha_1 + n_2\alpha_2 + n_3\alpha_3)\Big\},$$

*where* $\alpha_j = \left(\begin{smallmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{smallmatrix}\right)$.

3. *The representation of the space reflection $P(x^0, x) = (x^0, -x)$, where $x \in \mathbb{R}^3$, has the form*

$$g(P) = \eta_P \gamma^0, \quad \text{where } \eta_P = \pm i \ \text{ or } \ \eta_P = \pm 1.$$

4. *The operator of time reflection $T(x^0, x) = (-x^0, x)$ is represented as*

$$g(T) = \eta_T \gamma^0 \gamma^1 \gamma^3, \quad \text{where } |\eta_T| = 1.$$

We leave the proof of this lemma as an exercise.

In statements 3 and 4 we indicated the image of the representation within the multivalence adopted in the physical literature. Let us point out once more that the spinor representation into the group $\mathrm{SL}(4,\mathbb{C})$ is defined up to multiplication by the 4th roots of one, i.e., $\pm 1$ and $\pm i$.

The statement 1 of the lemma implies that under the natural decomposition of the space $\mathbb{C}^4$ into the sum $\mathbb{C}^2 \oplus \mathbb{C}^2$,

$$\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}, \quad \varphi \in \mathbb{C}^2, \quad \chi \in \mathbb{C}^2,$$

the restriction of the spinor representation of the group $\mathrm{O}(1,3)$ to the subgroup $\mathrm{SO}(3)$ breaks up into a sum of two isomorphic irreducible spinor representations described in 15.2.2.

In the space of spinors $\psi = \binom{\varphi}{\chi}$ we can introduce the new basis

$$\eta = \frac{\varphi + \chi}{\sqrt{2}}, \quad \xi = \frac{\varphi - \chi}{\sqrt{2}}.$$

The statements 1 to 3 of the lemma imply that the "semispinors" $\eta$ and $\xi$ transform independently under transformations in the proper Lorentz group $\mathrm{SO}(1,3) \subset \mathrm{O}(1,3)$ and go into each other under the space reflection:

$$g(P)\colon \ \eta \to \xi, \quad \xi \to \eta,$$
$$P(x^0, x) = (x^0, -x).$$

The actions of the group $\mathrm{SO}(1,3)$ on the semispinors $\eta$ and $\xi$ are called *semispinor representations* of the group $\mathrm{SO}(1,3)$. They are denoted by $g_+$ for $\eta$ and $g_-$ for $\xi$. The semispinor representations separately do not extend to the entire Lorentz group $\mathrm{O}(1,3)$.

Another description of the semispinor representation will be given in 15.3.2.

In physics semispinors are also called *Weyl's spinors*.

Note that in the basis $\psi = \binom{\eta}{\xi}$ the matrix $\gamma^0$ (the matrix of space reflection) has the form

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

(it mixes up semispinor representations).

The spinor representation of the group $\mathrm{O}(3,1)$ is not unitary, but the formula

$$\langle \psi, \psi \rangle = \psi^+ \gamma^0 \psi, \tag{15.31}$$

where $\psi^+ = (\overline{\psi_0}, \overline{\psi_1}, \overline{\psi_2}, \overline{\psi_3})$ is the row vector that is complex-conjugate to the column $\psi$, specifies an indefinite scalar product, which is invariant relative to the spinor representation.

The spinor

$$\bar{\psi} = \psi^+\gamma^0 = (\overline{\psi_0}, \overline{\psi_1}, -\overline{\psi_2}, -\overline{\psi_3})$$

is said to be *Dirac-conjugate* to $\psi$.

In terms of semispinors (in the basis $\psi = \binom{\eta}{\xi}$) the scalar product $\langle\psi, \psi\rangle = \bar{\psi}\psi$ has the form

$$\langle\psi, \psi\rangle = \xi^+\eta + \eta^+\xi.$$

Other quantities of the form $\bar{\psi}X\psi$, where $X \in \mathrm{M}(n, \mathbb{C})$, are transformed as tensors:

1) $\bar{\psi}\gamma^a\psi$ is a vector;
2) $\bar{\psi}\gamma^a\gamma^b\psi$ is a tensor of rank 2;
3) $\bar{\psi}\gamma^a\gamma^b\gamma^c\psi$ is a tensor of rank 3;
4) $\bar{\psi}\gamma^5\psi = \bar{\psi}\gamma^0\gamma^1\gamma^2\gamma^3\psi$ is a tensor of rank 4 (pseudoscalar).

One of the semispinor representations of the group $\mathrm{SO}(1, 3)$ is isomorphic to the standard representation of the group $\mathrm{SL}(2, \mathbb{C})$, and the other to the conjugate representation (see 15.3.2). This implies that the semispinor representations of the group $\mathrm{SO}(1, 3)$ have no nonzero invariant scalar products.

The isomorphism between the Lie algebras of the group $\mathrm{SL}(2, \mathbb{C})$ and the Lorentz group is a particular case of the isomorphism (11.2) for $n = 2$, where the Lie algebra of conformal (linear-fractional) transformations of the sphere $S^2$ is realized as the Lie algebra of the group $\mathrm{SO}(1, 3)$.

REMARK. If we select the generators $\tilde{\gamma}^0 = \gamma^0$, $\tilde{\gamma}^a = -i\gamma^a$ in $\mathrm{M}(4, \mathbb{C})$, we will obtain the relations

$$\{\tilde{\gamma}^a, \tilde{\gamma}^b\} = 2\delta^{ab} \cdot 1.$$

Using them, a spinor representation of the group $\mathrm{SO}(4)$ can be constructed along the same lines as before.

**15.2.4. Dirac equation.** The algebra of $\gamma$-matrices with relations (15.29) was introduced by Dirac for decomposition of the Klein–Gordon operator

$$\Box + m^2 = g^{ab}\,\frac{\partial}{\partial x^a}\,\frac{\partial}{\partial x^b} + m^2$$

(here $g_{ab}$ is the Minkowski metric and for simplicity we set $c = 1$) into a product of the first-order operators

$$-(\Box + m^2) = \left(i\gamma^a\,\frac{\partial}{\partial x^a} + m\right)\left(i\gamma^b\,\frac{\partial}{\partial x^b} + (-m)\right). \tag{15.32}$$

It is not hard to calculate that the decomposition (15.32) is equivalent to the relations

$$\{\gamma^a, \gamma^b\} = 2g^{ab} \cdot 1.$$

Therefore, if we take the coefficients to be the $\gamma$-matrices of the form (15.30), then such a representation will be possible.

The *Dirac equation* is the equation

$$\left(i\gamma^a \frac{\partial}{\partial x^a} - m\right)\psi = 0 \tag{15.33}$$

for the spinor field $\psi$. Here we set $c = 1$, where $c$ is the velocity of light in vacuum. Otherwise the mass would enter into the equation via the term $-mc^2\psi$.

If a field $\psi$ satisfies Dirac's equation, then the field

$$\bar{\psi} = \psi^+\gamma^0 = (\overline{\psi_0}, \overline{\psi_1}, -\overline{\psi_2}, -\overline{\psi_3})$$

(with entries of the row $\psi^+$ complex-conjugate to those of the column $\psi$) satisfies the equation

$$i\,\frac{\partial\bar{\psi}}{\partial x^a}\,\gamma^a + m\bar{\psi} = 0. \tag{15.34}$$

This can be shown by direct calculations.

The Dirac equation is the Euler–Lagrange equation for the action

$$S = \int \left[\frac{i}{2}\left(\bar{\psi}\gamma^a \frac{\partial\psi}{\partial x^a} - \frac{\partial\bar{\psi}}{\partial x^a}\gamma^a\psi\right) - m\bar{\psi}\psi\right] d^4x,$$

where $\psi$ and $\bar{\psi}$ are regarded as independent. The energy-momentum tensor and the current are

$$T^{ab} = \frac{i}{2}\, g^{ac}\left(\bar{\psi}\gamma^b \frac{\partial\psi}{\partial x^c} - \frac{\partial\bar{\psi}}{\partial x^c}\gamma^b\psi\right) = T^{ba},$$
$$J^a = \bar{\psi}\gamma^a\psi, \quad a = 0, 1, 2, 3.$$

The quantity $J^0 = \bar{\psi}\gamma^0\psi = \psi^+(\gamma^0)^2\psi = \psi^+\psi$ is called the *charge density*. Hence the charge itself is

$$Q = \int \psi^+\psi\, d^3x = \int J^0\, d^3x.$$

Obviously, the charge is nonnegative, $Q \geq 0$. At the same time the energy density $T^{00}$ is not positive definite, which leads to certain problems. However, we will not discuss them here.

For solutions to Dirac's equation of the type

$$\psi = \text{const} \cdot e^{i\langle k, x\rangle},$$

the wave vector lies on the mass surface

$$\langle k, k\rangle = m^2$$

(here $\hbar = 1$ and $c = 1$). If the momentum lies on the upper part of the mass surface ($k_0 > 0$), then such a solution is identified with a particle. The condition $k_0 > 0$ follows from the requirement that the energy of a particle is positive.

In the basis of semispinors $(\eta, \xi)$, the Dirac equation (15.33) becomes

$$i\,\frac{\partial \eta}{\partial t} = i\sum_{a=1}^{3} \sigma_a\,\frac{\partial \eta}{\partial x^a} + m\xi,$$
$$i\,\frac{\partial \xi}{\partial t} = -i\sum_{a=1}^{3} \sigma_a\,\frac{\partial \xi}{\partial x^a} + m\eta.$$

In the case of zero mass $m = 0$, Dirac's equation breaks up into two independent equations (*Weyl equations*) that describe particles of zero mass whose laws of motion are not invariant relative to spatial reflections (since the semispinor representation of the group $SO(1,3)$ is not extendable to the spatial reflection). Note that, as was pointed out before, the semispinor representation does not possess a nontrivial invariant scalar product, which could produce terms of mass type in the Lagrangian.

The inclusion of an exterior electromagnetic field is done according to the general rule,

$$p_a \to p_a + eA_a$$

(here again $\hbar = c = 1$). Hence we must make the following change in the Lagrangian:

$$\frac{\partial}{\partial x^a} \to \frac{\partial}{\partial x^a} - ieA_a(x),$$

where $e$ is the charge. Dirac's equation (15.33) and the conjugate equation (15.34) in the exterior field become

$$\begin{aligned} &i\gamma^a\Big(\frac{\partial}{\partial x^a} - ieA_a\Big)\psi - m\psi = 0,\\ &i\Big(\frac{\partial}{\partial x^a} + ieA_a\Big)\bar{\psi}\gamma^a + m\bar{\psi} = 0. \end{aligned} \tag{15.35}$$

Consider the matrix $C = -\tilde{\gamma}^2\tilde{\gamma}^0 = i\gamma^2\gamma^0$. It has the following properties:

$$C^{-1}\tilde{\gamma}^a C = -(\tilde{\gamma}^a)^{\top}, \quad C^{\top} = C^{-1} = -C \tag{15.36}$$

(here $\top$ means transposition). Using this matrix, define the operator of *charge conjugation*

$$\psi \to \psi^c = C\bar{\psi}^{\top}, \quad \bar{\psi} \to \bar{\psi}^c = \psi^{\top}C^{-1}.$$

Formula (15.36) and the Dirac equation imply the following theorem.

**Theorem 15.5.** *If the fields $\psi$ and $\bar{\psi}$ satisfy Dirac's equations* (15.35) *with charge $e$ in the field $A_a(x)$, then the fields $\psi^c$ and $\bar{\psi}^c$ satisfy equations* (15.35) *in the same field $A_a(x)$, but with charge of opposite sign, $e \to -e$.*

We see that the transformation of charge conjugation changes the sign of the charge of the particle described by the field $\psi$. This implies that the spinor representation and Dirac's equation describe two types of particles: particles with charge $e$ and those with charge $-e$. The same solution to Dirac's equations (15.35) determines the wave functions of an electron $\psi(x)$ and a positron $\psi^c(x)$.

**Lemma 15.8.** *The operator of charge conjugation*

1) *commutes with the group* $\mathrm{SO}(1,3)$;

2) *commutes with the spacial reflection $g(P) = \eta_P\gamma^0$ if $\eta_P = \pm i$, and anticommutes if $\eta_P = \pm 1$;*

3) *commutes with the time reflection $g(T)$ if $\eta_T = \pm 1$.*

We defined the operators of spacial and time conjugation in Lemma 15.7. The proof of Lemma 15.8 is left as an exercise.

**15.2.5. Clifford algebras.** The constructions of spinor representations of the groups $\mathrm{SO}(3)$ and $\mathrm{O}(1,3)$ are special cases of a more general construction, which we briefly describe for the real Clifford algebras.

Let $\langle u, v\rangle$ be a nondegenerate scalar product of type $(p,q)$ on $\mathbb{R}^{p,q}$. Consider the algebra generated by the $\mathbb{R}$-valued scalars and $\mathbb{R}^{p,q}$-valued vectors with usual addition and multiplication by a scalar $\lambda \in \mathbb{R}$ and the formal associative multiplication of vectors

$$u, v \to u \cdot v.$$

We will require that the identities

$$(u' + u'')v = u'v + u''v, \quad (\lambda u)v = \lambda(uv), \quad \lambda \in \mathbb{R},$$

and their analogs for the second factor $v$ be fulfilled. Moreover, we impose the additional condition

$$uv + vu = -2\langle u, v\rangle \in \mathbb{R}, \quad u, v \in \mathbb{R}^{p,q}.$$

The algebra thus obtained is called the *Clifford algebra* of a quadratic form of type $(p,q)$ and is denoted by $\mathrm{Cl}_{p,q}$.

If $e_1, \dots, e_n$ is a basis in $\mathbb{R}^{p,q}$, $n = p + q$, then the Clifford algebra is linearly generated by all products

$$v_{i_1} \cdots v_{i_k}, \quad i_1 < \cdots < i_k,$$

and is isomorphic as a linear space to the exterior algebra $\Lambda^*\mathbb{R}^n$. The isomorphism is specified by the mapping

$$v_{i_1}\cdots v_{i_k} \to v_{i_1}\wedge\cdots\wedge v_{i_k}.$$

Indeed, any monomial can be reduced to this form using the commutation rule $uv = -vu - 2\langle u, v\rangle$. It is natural to set the degree of such a monomial to be $\deg = k$.

Taking into account the multiplication, the Clifford algebras have a more complicated structure than the exterior algebras. For example, we have

$$\mathrm{Cl}_{1,0} = \mathbb{C}, \quad \mathrm{Cl}_{0,1} = \mathbb{R}\oplus\mathbb{R}, \quad \mathrm{Cl}_{2,0} = \mathbb{H}, \quad \mathrm{Cl}_{0,2} = \mathrm{Cl}_{1,1} = \mathrm{M}(2,\mathbb{R}).$$

The automorphism of the vector space $\alpha\colon v \to -v$ gives rise to an automorphism of the Clifford algebra, which breaks up into a sum of two subspaces formed by linear combinations of elements of even and odd degree, respectively,

$$\mathrm{Cl} = \mathrm{Cl}^0 + \mathrm{Cl}^1, \quad \alpha|_{\mathrm{Cl}^0} = 1, \quad \alpha|_{\mathrm{Cl}^1} = -1.$$

Consider the multiplicative group of all invertible elements $g$ of the Clifford algebra (which means that there exists an element $g^{-1}$ such that $gg^{-1} = g^{-1}g = 1$), and take its subgroup generated (relative to multiplication) by the vectors $v$ of unit length, $\langle v, v\rangle = 1$. This subgroup is denoted by $\mathrm{Pin}(p,q)$, and its intersection with elements of even degree is the group

$$\mathrm{Spin}(p,q) = \mathrm{Pin}(p,q)\cap\mathrm{Cl}^0.$$

These groups act on the vector space $\mathbb{R}^{p,q}$ by the formula

$$v \to gvg^{-1},$$

and this action preserves the form $\langle u, v\rangle$. Therefore, the following representations hold:

$$\mathrm{Pin}(p,q)\to\mathrm{O}(p,q), \quad \mathrm{Spin}(p,q)\to\mathrm{SO}(p,q).$$

Furthermore, the kernels of these homomorphisms consist of elements $\pm1$; by inversion of these homomorphisms we obtain the two-valued representations

$$\mathrm{O}(p,q)\to\mathrm{Pin}(p,q), \quad \mathrm{SO}(p,q)\to\mathrm{Spin}(p,q),$$

which are what was introduced as spinor representations.

## 15.3. Yang–Mills fields

**15.3.1. Gauge-invariant Lagrangians.** Given a Lagrangian

$$L\Big(\psi, \frac{\partial\psi}{\partial x^\alpha}\Big),$$

with $\psi$ a vector-function defined in a domain $D$ with coordinates $x^1, \dots, x^n$, suppose that this Lagrangian is invariant under the action of a matrix Lie group $G$ on vectors:

$$L\Big(\psi, \frac{\partial\psi}{\partial x^\alpha}\Big) = L\Big(g\psi, g\frac{\partial\psi}{\partial x^\alpha}\Big).$$

As the simplest example we may consider the massive scalar field

$$L = g^{\alpha\beta}\,\frac{\partial\bar\psi}{\partial x^\alpha}\,\frac{\partial\psi}{\partial x^\beta} - m^2\bar\psi\psi$$

described in 14.2.6 (we again set $\hbar = c = 1$) with the group $G = \{e^{i\varphi}\} = \mathrm{U}(1)$ acting by the formulas

$$\psi \to e^{i\varphi}\psi, \quad \bar\psi \to e^{-i\varphi}\bar\psi, \quad \varphi = \mathrm{const} \in \mathbb{R},$$
$$L\Big(e^{i\varphi}\psi, e^{-i\varphi}\bar\psi, e^{i\varphi}\frac{\partial\psi}{\partial x^\alpha}, e^{-i\varphi}\frac{\partial\bar\psi}{\partial x^\alpha}\Big) = L\Big(\psi, \bar\psi, \frac{\partial\psi}{\partial x^\alpha}, \frac{\partial\bar\psi}{\partial x^\alpha}\Big).$$

Now the question is how to construct a Lagrangian which would be invariant under local transformations

$$\psi(x) \to g(x)\psi(x)$$

by which the group $G$ acts independently at each point (the *locality principle*).

It turns out that this can be done if we replace the usual derivative by the covariant derivative:

$$\frac{\partial\psi}{\partial x^\alpha} \to \nabla_\alpha\psi = \frac{\partial\psi}{\partial x^\alpha} + A_\alpha\psi,$$
$$L\Big(\psi, \frac{\partial\psi}{\partial x^\alpha}\Big) \to L(\psi, \nabla_\alpha\psi),$$

where $A_\alpha$ is the gauge field with group $G$, and by definition

$$\nabla_\alpha(g(x)\psi(x)) = g(x)\nabla_\alpha\psi(x).$$

We recall the notion of the *gauge field* (*connection*), introduced in 10.1.3.

Let $D$ be a domain in $\mathbb{R}^n$, and let $G$ be a matrix Lie group acting on $k$-dimensional vectors:

$$\xi \to g\xi, \quad g \in G \subset \mathrm{GL}(k),\ \xi \in \mathbb{R}^k.$$

The covariant differentiation of functions $\psi\colon D \to \mathbb{R}^k$ is defined by the formula

$$\nabla_\alpha = \frac{\partial\psi}{\partial x^\alpha} + A_\alpha\psi,$$

where $A_\alpha(x)$ is a function on $D$ taking values in the set of $k \times k$ matrices lying in the Lie algebra of the group $G$ and $\alpha = 1, \dots, n$. Moreover, the covariant derivative is assumed to satisfy the following two conditions.

1. Under coordinate changes $y = y(x)$ in the domain $D$, the covariant derivative behaves as a covector:

$$\tilde{\nabla}_\mu \psi = \frac{\partial \psi}{\partial y^\mu} + \tilde{A}_\mu \psi = \frac{\partial x^\alpha}{\partial y^\mu}\Big(\frac{\partial \psi}{\partial x^\alpha} + A_\alpha \psi\Big) = \frac{\partial x^\alpha}{\partial y^\mu}\, \nabla_\alpha \psi,$$

which is equivalent to the following condition:

$$\tilde{A}_\mu = \frac{\partial x^\alpha}{\partial y^\mu}\, A_\alpha .$$

2. If we change the basis in $\mathbb{R}^k$ by means of a transformation from the group $G$, this change being dependent on the point $x$, $g = g(x)$, then

$$g(x)\nabla_\alpha \psi(x) = \nabla_\alpha (g(x)\psi(x)).$$

A sufficient condition for this relation to hold is that under *gauge transformations*

$$\psi \to g(x)\psi$$

the field $A_\alpha$ be also transformed by the rule

$$A_\alpha(x) \to g(x)A_\alpha(x)g(x)^{-1} - \frac{\partial g(x)}{\partial x^\alpha}\, g^{-1}(x) \tag{15.37}$$

(see Theorem 10.7).

The gauge field for the group $\mathrm{U}(1) = \{e^{i\varphi}\}$ is $A_\alpha = iB_\alpha$, where $B_\alpha$ is a real covector field, for which the transformation (15.37) has the form

$$\psi \to e^{i\varphi(x)}\psi, \quad B_\alpha \to B_\alpha - \frac{\partial \varphi}{\partial x^\alpha}.$$

Now we state the following general theorem.

**Theorem 15.6.** *Under transformations of the form*

$$\psi(x) \to g(x)\psi(x), \tag{15.38}$$

$$A_\alpha(x) \to g(x)A_\alpha(x)g^{-1}(x) - \frac{\partial g(x)}{\partial x^\alpha}\, g^{-1}(x), \tag{15.39}$$

*the covariant derivative*

$$\nabla_\alpha \psi = \frac{\partial \psi}{\partial x^\alpha} + A_\alpha \psi \tag{15.40}$$

*transforms by the rule*

$$\nabla_\alpha \psi \to g(x)\nabla_\alpha \psi. \tag{15.41}$$

*If the Lagrangian $L(\psi, \frac{\partial \psi}{\partial x^\alpha})$ is invariant under transformations $\psi \to g\psi$, $g = \text{const} \in G$:*

$$L\Big(\psi, \frac{\partial \psi}{\partial x^\alpha}\Big) = L\Big(g\psi, g\,\frac{\partial \psi}{\partial x^\alpha}\Big), \quad g = \text{const} \in G, \tag{15.42}$$

*then the Lagrangian $\tilde{L}(\psi, \frac{\partial \psi}{\partial x^\alpha}, A_\alpha) = L(\psi, \nabla_\alpha \psi)$ is invariant under transformations* (15.38) *and* (15.39).

**Proof.** Formulas (15.38), (15.39), and (15.40) imply that

$$\nabla_\alpha\psi(x) \to \frac{\partial}{\partial x^\alpha}(g(x)\psi) + \Big[gA_\alpha g^{-1} - \frac{\partial g(x)}{\partial x^\alpha}\, g^{-1}\Big] g\psi$$
$$= g(x)\Big[\frac{\partial\psi}{\partial x^\alpha} + A_\alpha\psi\Big] = g(x)\nabla_\alpha\psi.$$

The covariant derivative of the vector-function $\psi$ is a covector:

$$\nabla_\alpha\psi(x) = \nabla_\beta\psi(y)\,\frac{\partial y^\beta}{\partial x^\alpha}, \quad y = y(x).$$

Therefore, equalities (15.41) and (15.42) imply invariance of the new Lagrangian:

$$\tilde{L}(\psi, \nabla_\alpha\psi, A_\alpha) = \tilde{L}(g\psi, g\nabla_\alpha\psi, A'_\alpha), \quad A'_\alpha = gA_\alpha g^{-1} - \frac{\partial g}{\partial x^\alpha}\, g^{-1}.$$

Hence the theorem. □

In physics a gauge field $A_\alpha$ is also called a *compensating* field. In geometry it is known as a *connection* on a bundle. The gauge transformations (15.39) form a group, which is also called a *gauge group*.

Example. The following lemma, which is proved by straightforward calculations, provides an important example of a connection.

**Lemma 15.9.** *Each connection $A_\alpha$ with values in the algebra $\mathfrak{g}$ specifies also a covariant differentiation of vector fields with values in $\mathfrak{g}$ by the rule*

$$\nabla_\alpha B = \frac{\partial B}{\partial x^\alpha} + [A_\alpha, B], \tag{15.43}$$

*where $[A_\alpha, B] = A_\alpha B - BA_\alpha$ is the commutator in the Lie algebra. Under the gauge transformations*

$$B \to g(x)Bg^{-1}(x),$$
$$A_\alpha(x) \to g(x)A_\alpha(x)g^{-1}(x) - \frac{\partial g(x)}{\partial x^\alpha}\, g^{-1},$$

*the covariant derivative $\nabla_\alpha B$ transforms by the rule*

$$\nabla_\alpha B \to g(x)(\nabla_\alpha B)g^{-1}(x).$$

So far we dealt with covariant differentiation of vector-functions defined in a domain $D \in \mathbb{R}^n$. If we turn from a domain to a smooth manifold $M^n$, the notion of a vector-function must be extended as follows.

A manifold $E$ is called a (smooth) *vector bundle* over a manifold $M$ if its points are pairs $(x, \xi)$ consisting of a point $x \in M^n$ and a vector $\xi$ attached to this point, and the projection mapping

$$\pi\colon E \to M$$

is defined on $E$, which assigns to each point $(x, \xi)$ its "first coordinate", the point $x \in M^n$. Moreover, the following conditions are assumed.

1. The manifold $M$ is covered by domains $U_\alpha$ such that in each of them, the mapping $\pi$ is a projection of a Cartesian product to its first factor, i.e., every domain $\pi^{-1}(U_\alpha)$ is endowed with coordinates $x^1_\alpha, \ldots, x^n_\alpha$, $\xi^1_\alpha, \ldots, \xi^n_\alpha$ in which $\pi$ has the form

$$\pi(x, \xi) = x,$$

with $x^1_\alpha, \ldots, x^n_\alpha$ being coordinates in the domain $U_\alpha \subset M$.

2. On the intersections of any two such domains $U_\alpha \cap U_\beta$ there are changes of coordinates of the form

$$x_\alpha = x_\alpha(x_\beta), \quad \xi_\alpha = f_{\alpha\beta}(x)\xi_\beta,$$

where the functions $f_{\alpha\beta}(x)$ take values in the matrix Lie group $G$.

Here the following terminology is used: the manifold $E$ is called the *total space* of the vector bundle, the manifold $M$ is the *base* of the vector bundle, and the vector space $F = \mathbb{R}^k$, to which the vectors $\xi$ belong, is a *fiber* of the vector bundle.

The group $G$ is called the *structure group* of the bundle.

The function $\psi$, which assigns to each point of $M$ a vector $\xi$ attached to this point,

$$\psi \colon M \to E, \quad \pi(\psi(x)) = x,$$

is called a *section* of the vector bundle.

If the fibers of the bundle are complex vector spaces $\mathbb{C}^k$ and the group $G$ acts on them by linear transformations from $\mathrm{GL}(k, \mathbb{C})$, then the bundle is said to be *complex*.

If we want to differentiate sections of a bundle to obtain, as a result, other sections and to make the procedure independent of the choice of coordinates, then we must introduce a connection, i.e., a family of 1-forms with values in the Lie algebra $\mathfrak{g}$ of the group $G$, and put

$$\nabla_k \psi = \frac{\partial \psi}{\partial x^k} + A_k(x)\psi.$$

This formula will specify in each domain $U_\alpha$ covariant derivatives of the section $\psi$. These covariant derivatives will agree with gauge transformations $\xi \to g_{\alpha\beta}(x)\xi$.

EXAMPLES. 1. A Cartesian product $M \times \mathbb{R}^k$ is called a trivial bundle.

2. The manifold $TM^n$ of all pairs $(x, \xi)$, where $x \in M^n$ and $\xi$ is a tangent vector at the point $x$, is called the *tangent bundle*.

**15.3.2. Covariant differentiation of spinors.** Let $M^4$ be a manifold with metric $g_{ab}$ of type $(1,3)$. A connection compatible with this metric generates a covariant differentiation of spinors.

The transformation

$$\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix} \to \begin{pmatrix} \eta \\ \xi \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi + \chi \\ \varphi - \chi \end{pmatrix}$$

gives rise to the *semispinors* $\eta$ and $\xi$, and the spinor representation of the group $\mathrm{SO}(1,3)$ breaks up into a Cartesian sum of conjugate semispinor representations

$$\eta \to g\eta, \quad \xi \to \bar{g}\xi, \quad g \in \mathrm{SL}(2,\mathbb{C}), \tag{15.44}$$

on the space $\mathbb{C}^2$, which is called the space of *two-component spinors* (or *semispinors*) (see 15.2.3).

In order to describe this representation $\mathrm{SO}(1,3) \to \mathrm{SL}(2,\mathbb{C})$, we represent the Minkowski space as the space of Hermitian $2 \times 2$ matrices

$$Y = \begin{pmatrix} y^0 + y^3 & y^1 + iy^2 \\ y^1 - iy^2 & y^0 - y^3 \end{pmatrix} = y^k \sigma_k, \quad \bar{Y}^\top = Y,$$

where $\sigma_0$ is the identity matrix, and $\sigma_1$, $\sigma_2$, $\sigma_3$ are the Pauli matrices. The scalar product is specified by the determinant:

$$\langle Y, Y \rangle = \det Y = (y^0)^2 - \sum_{k=1}^{3} (y^k)^2.$$

The transformations

$$Y \to gY\bar{g}^\top, \quad g \in \mathrm{SL}(2,\mathbb{C}),$$

preserve the Minkowski metric and hence determine a representation

$$\mathrm{SL}(2,\mathbb{C}) \to \mathrm{SO}(1,3),$$

which takes the matrices $\pm A$ into the same element of $\mathrm{SO}(1,3)$. The inverse two-valued mapping

$$\mathrm{SO}(1,3) \to \mathrm{SL}(2,\mathbb{C})$$

is the representation of the group $\mathrm{SO}(1,3)$ on the space of semispinors.

In order to define covariant differentiation of spinors, it suffices to define that of semispinors.

The covariant derivative of a semispinor field is defined as

$$\widehat{\nabla}_\alpha \eta = \frac{\partial \eta}{\partial x^\alpha} + A_\alpha \eta, \tag{15.45}$$

where $A_\alpha$ is an element of the Lie algebra of the group $\mathrm{SL}(2,\mathbb{C})$. Conjugate semispinors are differentiated by the rule

$$\widehat{\nabla}_\alpha \xi = \frac{\partial \xi}{\partial x^\alpha} + \overline{A}_\alpha \xi. \tag{15.46}$$

We must define the connection $A_\alpha \, dx^\alpha$.

To this end, at each point of the curved space $M^4$ we choose a collection of matrices $\gamma^a$ satisfying the conditions $\gamma^a\gamma^b + \gamma^b\gamma^a = g'^{ab} \cdot 1$, where $g'^{ab}$ is the Minkowski metric. The matrices $\sigma_k$ are chosen in a similar manner. They depend on the point of the space, but they must satisfy the same commutation conditions: the metric on the space of Hermitian $2\times 2$ matrices at each point of the space has the form $\langle \sigma_a, \sigma_b\rangle = g'_{ab}$. Since the space of Hermitian matrices at each point is identified in a natural way with the tangent space at the point, the vectors $\sigma_a$ must form a tetrad at each point:

$$\langle \sigma_a, \sigma_b\rangle = \begin{cases} 1 & \text{for } a = b = 0, \\ -1 & \text{for } a = b = 1, 2, 3, \\ 0 & \text{for } a \neq b. \end{cases}$$

Since we identify the tangent vectors with Hermitian matrices, we take the usual symmetric connection $\nabla$ compatible with the metric on the manifold $M^n$ and define covariant differentiation of Hermitian matrices as that of tangent vectors:

$$\hat{\nabla}_a Y = \sigma_b (\nabla_a y)^b, \quad Y = y^a \sigma_a.$$

The components $\Gamma^c_{ab}$ of the connection in terms of tetrads were computed in 10.2.5.

Each local transformation in $\mathrm{SO}(1,3)$ generates, by means of the spinor representation, a local transformation of spinors. Therefore, the connection on the tangent bundle $TM^n$ compatible with metric generates a connection on spinor fields: the matrices of the algebra $o(1,3)$ specifying the connection on $TM^4$ are mapped homomorphically into matrices of $\mathfrak{sl}(4)$ specifying the connection on spinor fields. In terms of semispinors this mapping looks particularly simple.

**Lemma 15.10.** *In terms of semispinors (see formulas* (15.45) *and* (15.46)) *the covariant derivative of spinor fields has the form*

$$A_a = -\frac{1}{2}\Gamma^c_{ba}\sigma_c\sigma_b,$$

*where in summation over $b$ the sign is taken into account: we assign "plus" to the term corresponding to $b = 0$ and "minus" to the other terms.*

The proof of this lemma is left as a problem to the reader.

The inclusion of an electromagnetic field is done in the usual manner.

**15.3.3. Curvature of a connection.** The commutator of covariant derivatives in different directions is

$$[\nabla_\mu, \nabla_\nu]\psi = \nabla_\mu \nabla_\nu \psi - \nabla_\nu \nabla_\mu \psi = \frac{\partial}{\partial x^\mu}\Big(\frac{\partial \psi}{\partial x^\nu} + A_\nu \psi\Big)$$
$$+ A_\mu\Big(\frac{\partial \psi}{\partial x^\nu} + A_\nu \psi\Big) - \frac{\partial}{\partial x^\nu}\Big(\frac{\partial \psi}{\partial x^\mu} + A_\mu \psi\Big) - A_\nu\Big(\frac{\partial \psi}{\partial x^\mu} + A_\mu \psi\Big)$$
$$= \Big(\frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu]\Big)\psi.$$

Hence the commutator of the operators $\nabla_\mu$, $\nabla_\nu$ is the operator of multiplication by the matrix $F_{\mu\nu}$ of the form

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu]. \tag{15.47}$$

**Theorem 15.7.** *Under gauge transformations*

$$A_\alpha(x) \to g(x)A_\alpha(x)g^{-1}(x) - \frac{\partial g(x)}{\partial x^\alpha}\, g^{-1},$$

*the matrices $F_{\mu\nu}$ as in* (15.47) *transform as follows:*

$$F_{\mu\nu} \to gF_{\mu\nu}g^{-1}.$$

*The matrices $F_{\mu\nu}$ form an antisymmetric tensor of rank* 2 *with values in the Lie algebra $\mathfrak{g}$, i.e., under changes of coordinates $y = y(x)$ they transform by the rule*

$$F_{\mu\nu}(x) = F_{\alpha\beta}(y)\, \frac{\partial y^\alpha}{\partial x^\mu}\, \frac{\partial y^\beta}{\partial x^\nu}, \quad y = y(x).$$

The *proof* consists in a direct substitution.

The form

$$\Omega = \sum_{\mu<\nu} F_{\mu\nu}\, dx^\mu \wedge dx^\nu$$

with values in the Lie algebra $\mathfrak{g}$ is called the *curvature form* of the connection $A_\mu$.

A connection $A_\mu$ is said to be *trivial* if there exists a function $g_0(x)$ taking values in the group $G$ and such that

$$A_\mu(x) = -\frac{\partial g_0(x)}{\partial x^\mu}\, g_0^{-1}(x).$$

In this case the gauge transformation specified by the function $g_0(x)^{-1}$ reduces covariant differentiation to the form

$$\nabla_\alpha \psi = \frac{\partial \psi}{\partial x^\alpha}, \quad \tilde{A}_\alpha \equiv 0.$$

The curvature of this connection equals zero, and since the quantities $F_{\mu\nu}$ form a tensor, we obtain the proof of the following theorem.

**Theorem 15.8.** *The curvature form of a trivial connection is equal to zero.*

The converse is also true: is $F_{\mu\nu} \equiv 0$, then the connection is (locally) trivial.

Using the connection (15.43), we define the covariant differential of the form $\Omega$:

$$D\Omega = \sum_{\lambda<\mu<\nu} (\nabla_\lambda F_{\mu\nu} + \nabla_\nu F_{\lambda\mu} + \nabla_\mu F_{\nu\lambda})\, dx^\lambda \wedge dx^\mu \wedge dx^\nu.$$

It satisfies the *Bianchi identity*

$$D\Omega = 0. \tag{15.48}$$

EXAMPLE. Let the $\Gamma^i_{jk}$ be Christoffel symbols specifying covariant differentiation of tangent vector fields on an $n$-dimensional manifold:

$$\nabla_k \psi = \frac{\partial \psi}{\partial x^k} + A_k \psi, \quad A_k = \Gamma^i_{jk} \in \mathfrak{gl}(n).$$

Then the curvature form is

$$F_{ij} = R^l_{kij}, \quad F(X,Y)Z = R(Y,X)Z,$$

where $R^l_{kij}$ is the Riemann curvature tensor (see 10.2.1).

If a symmetric connection is compatible with metric, then we will specify the connection of terms of "tetrads" of vectors $e_1, \dots, e_n$ which form orthonormal bases in tangent spaces, $\langle e_i, e_j \rangle = \delta_{ij}$. The covariant derivatives for such a connection were calculated in 10.2.5. The connection has the form

$$(A_i)_{kj} = \Gamma_{kji}, \quad \nabla_{e_i} e_j = \sum_k \Gamma_{kji} e_k,$$

and the Christoffel symbols $\Gamma_{kji}$ are skew-symmetric in subscripts $k, j$ (i.e., the matrices $A_i$ belong to the Lie algebra of the group $\mathrm{SO}(n)$). The curvature form

$$(F_{ij})_{lk} = R_{lkij} = -R_{klij}$$

also takes values in this Lie algebra.

For a metric on a two-dimensional surface the curvature form is

$$F = \begin{pmatrix} 0 & \Omega \\ -\Omega & 0 \end{pmatrix}, \quad \Omega = K\sqrt{g}\, dx^1 \wedge dx^2,$$

where $K$ is the Gaussian curvature of the surface.

**15.3.4. The Yang–Mills equations.** As was shown in 15.3.1, the requirement that the Lagrangian is invariant relative to local transformations $\psi \to g(x)\psi$ determines the Lagrangian of interaction of the field $\psi$ with the gauge field $A_\alpha$. But this requirement does not specify the Lagrangian of the gauge field itself.

The Lagrangian $L = L(A_\mu, \partial A_\mu/\partial x^\nu)$ of a gauge field must possess the following properties: it must be a scalar and remain unchanged under gauge transformations.

The simplest Lagrangian satisfying these conditions is

$$L = -\frac{1}{4}\, g^{\mu\lambda} g^{\nu\kappa} \langle F_{\mu\nu}, F_{\lambda\kappa}\rangle, \tag{15.49}$$

where

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu],$$

$g_{\mu\nu} = g_{\mu\nu}(x)$ is the metric on the manifold, and $\langle\,,\rangle$ is the Killing form on the Lie group $G$.

Obviously, it is a scalar, and, since the curvature form transforms under gauge transformations by the rule

$$F_{\mu\nu} \to g F_{\mu\nu} g^{-1}$$

and the Killing metric is bilaterally invariant, we have

$$\langle g F_{\mu\nu} g^{-1}, g F_{\lambda\kappa} g^{-1}\rangle = \langle F_{\mu\nu}, F_{\lambda\kappa}\rangle.$$

Suppose that the metric $g_{\mu\nu}$ is Euclidean or pseudo-Euclidean, $g_{\mu\nu} = \varepsilon_\mu \delta_{\mu\nu}$, $\varepsilon_\mu = \pm 1$. Then the action functional becomes

$$S[A_\mu] = -\int \frac{1}{4}\, \langle F_{\mu\nu}, F_{\mu\nu}\rangle\, d^n x \tag{15.50}$$

(here summation over the repeated indices $\mu$, $\nu$ takes into account the signs of $\varepsilon_\nu, \varepsilon_\mu = \pm 1$).

**Theorem 15.9.** *Suppose the Killing form on a Lie group is nondegenerate. Then the Euler–Lagrange equations for extremals of the functional* (15.50) *are*

$$\nabla_\mu F_{\mu\nu} = 0, \tag{15.51}$$

*where*

$$\nabla_\mu F_{\mu\nu} \equiv \frac{\partial F_{\mu\nu}}{\partial x^\mu} + [A_\mu, F_{\mu\nu}]$$

(*with summation over* $\mu$).

**Proof.** For a small local variation $\delta A_\mu$ we have

$$\delta F_{\mu\nu} = \frac{\partial}{\partial x^\mu}\,\delta A_\nu - \frac{\partial}{\partial x^\nu}\,\delta A_\mu + [\delta A_\mu, A_\nu] + [A_\mu, \delta A_\nu],$$

$$\delta S = -\frac{1}{2}\int \langle F_{\mu\nu}, \delta F_{\mu\nu}\rangle\, d^n x.$$

Integrating by parts we prove the equality

$$\int \Big\langle F_{\mu\nu}, \frac{\partial}{\partial x^\mu}\,\delta A_\nu \Big\rangle\, d^n x = -\int \Big\langle \frac{\partial F_{\mu\nu}}{\partial x^\mu}, \delta A_\nu \Big\rangle\, d^n x.$$

Invariance of the Killing form implies that

$$\langle F_{\mu\nu}, [\delta A_\mu, A_\nu]\rangle = -\langle [F_{\mu\nu}, A_\nu], \delta A_\mu\rangle.$$

Substituting these two equalities into the expressions for $\delta S$ we obtain

$$\delta S = \frac{1}{2}\int \Big\{ \Big\langle \frac{\partial F_{\mu\nu}}{\partial x^\mu}, \delta A_\nu \Big\rangle - \Big\langle \frac{\partial F_{\mu\nu}}{\partial x^\nu}, \delta A_\mu \Big\rangle + \langle [F_{\mu\nu}, A_\nu], \delta A_\mu\rangle - \langle [F_{\mu\nu}, A_\mu], \delta A_\nu\rangle \Big\}\, d^n x.$$

After changing notation for the indices, we arrive at the equality

$$\delta S = \int \Big\langle \frac{\partial F_{\mu\nu}}{\partial x^\mu} + [A_\mu, F_{\mu\nu}], \delta A_\nu \Big\rangle\, d^n x = \int \langle \nabla_\mu F_{\mu\nu}, \delta A_\nu\rangle\, d^n x.$$

We see that $\delta S$ vanishes on extremals for an arbitrary variation $\delta A_\nu$. If the Killing form is nondegenerate, this means that $\nabla_\mu F_{\mu\nu} = 0$. Hence the theorem. □

Equations (15.51) are called the *Yang–Mills equations*.

Their solutions for the group $G = \mathrm{SU}(2)$ are called the *Yang–Mills fields*, although this term is often used in the case of a general group $G$.

EXAMPLES. 1. For the group $G = \mathrm{U}(1)$, the connection is an electromagnetic field, and the Lagrangian has the form

$$L = -\frac{1}{4} F_{\mu\nu}^2 = -\frac{1}{4}\Big(\frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}\Big)^2.$$

In this case equations (15.51) are the *Maxwell equations*

$$\frac{\partial F_{\mu\nu}}{\partial x^\mu} = 0$$

(see 14.2.6).

2. The Yang–Mills equations can be formally written for a group with a degenerate Killing form as well. In this case they, of course, cannot be derived from variational principles. For example, for a connection with values in the affine group (Cartan's connection; see 10.1.4) the Yang–Mills equations coincide with the Einstein equations

$$R_{ab} - \frac{1}{2} R g_{ab} = 0.$$

If the space is four-dimensional, then the action

$$S[A] = \int \mathrm{Tr}\, F_{\mu\nu}^2\, d^4 x$$

is invariant under conformal transformations. Indeed, the differential of any conformal transformation of $\mathbb{R}^4$ or $\mathbb{R}^{1,3}$ reduces to a dilatation and rotation (see Section 11.1). Since $F_{\mu\nu}$ is a tensor, for a dilatation $x \to \lambda x$ we have

$$F_{\mu\nu} \to \lambda^{-2} F_{\mu\nu}, \quad d^4x \to \lambda^4 \, d^4x.$$

Obviously,

$$\langle F_{\mu\nu}, F_{\mu\nu}\rangle \, d^4x \to \langle F_{\mu\nu}, F_{\mu\nu}\rangle \, d^4x.$$

Thus we have proved the following theorem.

**Theorem 15.10.** *The action* (15.50) *and the Maxwell–Yang–Mills equations* (15.51) *in the Minkowski space are conformally invariant, i.e., they have the group of symmetries* O(2, 4).

**15.3.5. Characteristic classes.** The curvature form allows for construction of important scalar-valued gauge-invariant forms.

Suppose that $A_\alpha \, dx^\alpha$ is a connection in a vector bundle, and let $\Omega = \sum_{\mu<\nu} F_{\mu\nu} \, dx^\mu \wedge dx^\nu$ be its curvature form. Consider the form

$$\tilde{c}_1 = \operatorname{Tr} \Omega = \sum_{\mu<\nu} \operatorname{Tr} F_{\mu\nu} \, dx^\mu \wedge dx^\nu.$$

Since

$$\operatorname{Tr} \Omega \to \operatorname{Tr}(g\Omega g^{-1}) = \operatorname{Tr} \Omega$$

under gauge transformations, this form is gauge-invariant. Furthermore, it is closed,

$$d\tilde{c}_1 = d \operatorname{Tr} \Omega = 0,$$

since locally it is an exact form, $\operatorname{Tr} \Omega = d \operatorname{Tr} A$. This can be shown by the following calculation:

$$\begin{aligned}
\operatorname{Tr} \Omega &= \sum_{\mu<\nu} \operatorname{Tr}\Big(\frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} + [A_\mu, A_\nu]\Big) \, dx^\mu \wedge dx^\nu \\
&= \sum_{\mu<\nu} \operatorname{Tr}\Big(\frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}\Big) \, dx^\mu \wedge dx^\nu = d \operatorname{Tr} A,
\end{aligned}$$

since the trace of a matrix commutator is equal to zero.

Now we write down the Bianchi identity (15.48),

$$\sum_{\lambda<\mu<\nu} (\nabla_\lambda F_{\mu\nu} + \nabla_\nu F_{\lambda\mu} + \nabla_\mu F_{\nu\lambda}) \, dx^\lambda \wedge dx^\mu \wedge dx^\nu = 0,$$

for the connection (15.43):

$$\nabla_\alpha B = \frac{\partial B}{\partial x^\alpha} + [A_\alpha, B].$$

Obviously, it can be rewritten in the form of the *Maurer–Cartan equation*

$$D\Omega = d\Omega + [A, \Omega] = 0,$$

where $A = A_\alpha \, dx^\alpha$ is a (locally specified) connection.

Now we define $2k$-forms

$$\tilde{c}_k = \operatorname{Tr}\Omega^k = \operatorname{Tr}(\Omega \wedge \cdots \wedge \Omega).$$

**Lemma 15.11.** *The forms $\tilde{c}_k$ are closed.*

**Proof.** We have

$$d\tilde{c}_k = d\operatorname{Tr}\Omega^k = \sum_{i=1}^{k} \operatorname{Tr}(\Omega^{i-1} \wedge d\Omega \wedge \Omega^{k-i}) = -\sum_{i=1}^{k} \operatorname{Tr}(\Omega^{i-1} \wedge [A, \Omega] \wedge \Omega^{k-i}).$$

As before, the equality $\operatorname{Tr}[A, \Omega] = 0$ implies that the traces of all terms vanish (we omit these straightforward calculations). Therefore, $d\tilde{c}_k = 0$. Hence the lemma. □

These closed forms posses an important property: their integrals over manifolds remain unchanged under variations of the connection.

**Theorem 15.11.** *The variational derivative of the functionals*

$$S_k[A_\mu] = \int_{M^{2k}} \tilde{c}_{2k} = \int_{M^{2k}} \operatorname{Tr}\Omega^k$$

*is identically equal to zero:*

$$\delta S_k[A_\mu] = \int_{M^{2k}} d(\operatorname{Tr}\delta A) = 0.$$

**Proof.** For brevity, we restrict ourselves to the proof for $k = 1$.

Under local variation of the connection $A_\mu \to A_\mu + \delta A_\mu$ the form $\operatorname{Tr}\Omega$ goes into $\operatorname{Tr}\Omega + \operatorname{Tr}\delta\Omega$. Note that the difference of two connections $\delta A$ is a covector, which is a 1-form with values in the Lie algebra. Hence we conclude that the form $\delta A$ is well defined and its exterior differential equals $\operatorname{Tr}\delta\Omega = d\operatorname{Tr}\delta A$.

Since the variation is local, we can apply the Stokes theorem to obtain

$$\delta S = \int_{M^2} \operatorname{Tr}\delta\Omega = \int_{M^2} d(\operatorname{Tr}\delta A) = 0.$$

Hence the theorem. □

A closed gauge-invariant form $\omega$ for which the functional $S = \int \omega$ has identically vanishing variational derivative,

$$\frac{\delta \int \omega}{\delta A} \equiv 0,$$

is called a differential-geometric *characteristic class.*

For the group $G = \mathrm{SO}(n)$, the Lie algebra consists of skew-symmetric matrices $A = -A^\top$. This implies that

$$(\Omega^{2k+1})^\top = (-1)^{2k+1}\Omega^{2k+1}$$

and the trace of this form is equal to zero:

$$\tilde{c}_{2k+1} = \mathrm{Tr}(\Omega^{2k+1}) = 0.$$

Hence for the group $G = \mathrm{SO}(n)$ the only nontrivial classes are

$$\tilde{p}_k = \tilde{c}_{2k}.$$

For the group $\mathrm{SO}(2n)$ there is one more characteristic class $\tilde{\chi}_n$. To define it, we recall that the *Pfaffian* of a skew-symmetric $2n \times 2n$ matrix $A = (a_{kl})$, $a_{kl} = -a_{lk}$, is given by the expression

$$\mathrm{Pf}(A) = \frac{1}{n!}\varepsilon^{i_1 \dots i_{2n}} a_{i_1 i_2} \wedge a_{i_3 i_4} \wedge \cdots \wedge a_{i_{2n-1} i_{2n}},$$

where $\varepsilon^{i_1 \dots i_{2n}}$ denotes the permutation $\left(\begin{smallmatrix} 1 & \cdots & 2n \\ i_1 & \cdots & i_{2n} \end{smallmatrix}\right)$ (in the classical setting, where the entries of $A$ are scalars, the formula involves usual multiplication). The characteristic class $\tilde{\chi}_n$ is defined as

$$\tilde{\chi}_n = \mathrm{Pf}(\Omega),$$

where $\Omega_{ij}$ is the curvature form,

$$\Omega = \sum_{\mu<\nu} F_{\mu\nu}\, dx^\mu \wedge dx^\nu.$$

For the groups $\mathrm{SO}(n)$ the classes $\tilde{p}_k$ are polynomials in the *Pontryagin characteristic classes* $p_k$, which are of degree $4k$ and are uniquely representable as polynomials in classes $\tilde{p}_i$.

The class $\tilde{\chi}_n$ is called the *Euler characteristic class.* More precisely, the Euler class is the class $\chi_n = \lambda_n \tilde{\chi}_n$, $\lambda_n = \mathrm{const}$, whose integral over a manifold equals its Euler characteristic. The Pontryagin classes are also integer-valued: their integrals over any cycle are integers.

Polynomials in characteristic classes are again characteristic classes.

A generic characteristic class for the group $\mathrm{SO}(2n)$ is a polynomial in $\tilde{\chi}_n, \tilde{p}_1, \dots, \tilde{p}_{n-1}$, and for the group $\mathrm{SO}(2n+1)$, a polynomial in $\tilde{p}_1, \dots, \tilde{p}_n$.

For the groups $G = \mathrm{U}(n)$ we obtain the *Chern characteristic classes* $c_k$, which are expressible in a one-to-one way as polynomials in $\tilde{c}_k$. These classes are also integer-valued and, for example,

$$c_1 = \frac{1}{2\pi}\,\tilde{c}_1.$$

The Lie algebra of the group $\mathrm{SU}(n)$ consists of zero-trace matrices; hence for $G = \mathrm{SU}(n) \subset \mathrm{U}(n)$, the first Chern class is always equal to zero,

$$\tilde{c}_1 = \mathrm{Tr}\,\Omega = 0.$$

EXAMPLE. For $n = 1$ the Euler class $\tilde{\chi}_1$ is

$$\tilde{\chi}_1 = \Omega_{12} = K\sqrt{g}\, dx^1 \wedge dx^2.$$

The variational derivative with respect to metric of the functional

$$S[g_{ij}] = \int_{\mathbb{R}^2} K\sqrt{g}\, dx^1 \wedge dx^2 = \int_{\mathbb{R}^2} \frac{1}{2} R\sqrt{g}\, dx_1 \wedge dx^2$$

is identically equal to zero (Theorem 14.7).

We will state without proof the following formulas for the Pontryagin, Chern, and Euler classes:

$$p = \det\left(1 + \frac{\Omega}{2\pi}\right) = 1 + p_1 + p_2 + \cdots ,$$
$$c = \det\left(1 + \frac{\Omega}{2\pi i}\right) = 1 + c_1 + c_2 + \cdots ,$$
$$\chi_n = \frac{\tilde{\chi}_n}{(2\pi)^n}.$$

The expressions of type $\det(1 + A)$, where $A$ is a 2-form with values in a matrix algebra, are calculated by formal rules of linear algebra. Note that the matrix entries commute (being 2-forms). As a result, we obtain a decomposition of such a determinant into a sum of forms which realize the integer-valued Pontryagin $p_k$ and Chern $c_k$ classes.

These matrix polynomials are invariant in the sense that they do not depend on gauge transformations which carry the curvature form $\Omega$ into $g\Omega g^{-1}$. Indeed,

$$\det(1 + \lambda g\Omega g^{-1}) = \det[g(1 + \lambda\Omega)g^{-1}] = \det(1 + \lambda\Omega).$$

For the Chern classes the coefficients of the form $\Omega$ belong to the algebra $\mathfrak{u}(n)$ which consists of matrices satisfying the condition $A^{\top} = -\bar{A}$. Their traces are purely imaginary, and on division of $\Omega$ by $2\pi i$ we obtain a form with real-valued matrix diagonal elements.

If the structure group of a complex bundle $\xi$ is $\mathrm{SU}(n)$, then $\operatorname{Tr} A = 0$ for any matrix in $\mathfrak{su}(n)$, and we have

$$c_1(\xi) = 0.$$

We leave it as a problem for the reader to prove that the classes $p_k$ and $c_l$ are expressible as polynomials in classes $\tilde{p}_k$ and $\tilde{c}_l$, respectively.

**15.3.6. Instantons.** There are a number of important cases where the problem of finding the global minima (called the *instantons*) of a functional $S[A]$ can be solved with the aid of the following trick.

Suppose that the functional $S[A]$ is representable as

$$S[A] = \int L(A) = \int L'(A) + \int \omega,$$

where $L'(A) \geq 0$ and $\omega$ is a characteristic class (the integral of $\omega$ does not change under variations of the field $A$). If there exists a field $A$ such that $L'(A) \equiv 0$, then it affords the global minimum to the functional $S$ among all the fields with a given value of the integral $\int \omega$.

There are important cases (one of which will be discussed below) where the condition $L' \equiv 0$ reduces to an equation of lower order than that of the Euler–Lagrange equation for the functional $S$. On the other hand, the problem $L' \equiv 0$ may have no solution (in which case the trick does not work).

Actually we have already applied this trick when describing harmonic mappings of $S^2$ into $S^2$ (see 14.2.5). In that case the quantity $\int \omega$ was the degree of the mapping.

Now we will give another simple example where this trick does not work.

Let $M^2$ be a closed oriented surface. For any smooth immersion of it in $\mathbb{R}^3$, we define the functional

$$W = \int_{M^2} H^2 \, d\mu,$$

where $H = \frac{k_1+k_2}{2}$ is the mean curvature of the surface and $d\mu = \sqrt{g}\, dx^1 \wedge dx^2$ is the area form in the induced metric. In physics such a functional arises for surfaces in multidimensional spaces and has the form

$$\int_{M^2} |H|^2 \, d\mu,$$

where $H$ is the mean curvature vector.

We represent the functional $W$ as

$$W = \int_{M^2} \Big(\frac{k_1 - k_2}{2}\Big)^2 \, d\mu + \int_{M^2} k_1 k_2 \, d\mu.$$

Since $K = k_1 k_2$ is the Gaussian curvature, we see that $K\, d\mu$ is a characteristic class:

$$\int_{M^2} k_1 k_2 \, d\mu = \text{const},$$

with the constant depending only on the topology of $M^2$ (by the Gauss–Bonnet theorem, it equals $2\pi\chi(M^2)$, where $\chi(M^2)$ is Euler's characteristic of the manifold $M^2$). The functional

$$W' = \int_{M^2} \Big(\frac{k_1 - k_2}{2}\Big)^2 \, d\mu$$

is called the *conformal area* or the *Willmore functional*. This functional is invariant under conformal transformations of the ambient space, and hence its extremals are always degenerate.

If $M^2$ is a sphere, then the global minimum of $W$ can be easily found as the solution to the equation $W' = 0$: these are circular spheres all of whose points are umbilical, i.e., $k_1 = k_2$. If $M^2$ is a torus, then this reasoning does not work: as we know, a closed surface with all points umbilical is homeomorphic to the sphere (see 4.4.3).

In mathematical physics there are a number of problems where the instanton trick is inapplicable. We now consider a setting where it works well.

Let $M^4$ be the four-dimensional sphere, which will be regarded as the space $\mathbb{R}^4$ with coordinates $x^0, x^1, x^2, x^3$ completed with a point at infinity. In this case we must impose the condition that in these coordinates $F_{ab} \to 0$ as $|x| \to \infty$, since the connection is extended to the sphere. This means that the connection converges to the trivial one:

$$A_a \sim -\frac{\partial g}{\partial x^a} g^{-1} \qquad \text{for} \quad |x| \to \infty.$$

The mapping $g(x)$ has different limits $g_\infty(v)$ along the rays $v|x|$, where $v$ is a vector on the unit sphere. Therefore, we have the mapping

$$g_\infty \colon S^3 \to \mathrm{SU}(2) = S^3.$$

Consider gauge fields $A_\alpha$ with the group $G = \mathrm{SU}(2)$. On a four-dimensional manifold the functional for the Yang–Mills fields is conformally invariant, hence the action with Lagrangian (15.49) can be written in Euclidean coordinates on $\mathbb{R}^4$ ($g_{ab} = \delta_{ab}$):

$$S[F] = -\int_{\mathbb{R}^4} \mathrm{Tr}(F_{ab}F_{ab})\, d^4x,$$

with summation over $a, b = 1, \dots, 4$ and $\Omega = \sum_{a<b} F_{ab}\, dx^a \wedge dx^b$ being the connection's curvature form. Note that $\mathrm{Tr}\, X^2 < 0$ for nonzero matrices in the Lie algebra of the group SU(2). Therefore, this functional is positive definite. With each 2-form $\Omega$ we associate the dual form

$$*\Omega = \frac{1}{2}\, \varepsilon_{abcd} F_{ab}\, dx^c \wedge dx^d$$

(with summation over all indices). It may be written as

$$\sum_{a<b} (*F)_{ab}\, dx^a \wedge dx^b = \sum_{c<d} F_{cd} *(dx^c \wedge dx^d),$$

where the right-hand side contains the usual Hodge operator on the four-dimensional Riemannian manifold.

We recall that the form $\mathrm{Tr}(\Omega \wedge \Omega)$ is a characteristic class whose integral is

$$D[F] = \int_{\mathbb{R}^4} \mathrm{Tr}\, F_{ab}F_{cd}\varepsilon^{abcd}\, d^4x = 2\int_{\mathbb{R}^4} \mathrm{Tr}\, F_{ab}(*F)^{ab}\, d^4x,$$

and it does not change under variations of the metric. It can be shown that the value of this functional is actually equal to

$$D = 4\pi^2 \deg g_\infty.$$

Now consider the functional

$$\begin{aligned}
T[F] &= -\frac{1}{2}\int_{\mathbb{R}^4} \operatorname{Tr}(F_{ab} - (*F)_{ab})(F^{ab} - (*F)^{ab})\, d^4x \\
&= -\frac{1}{2}\int_{\mathbb{R}^4} \operatorname{Tr} F_{ab}F^{ab}\, d^4x \\
&\quad -\frac{1}{2}\int_{|R^4} \operatorname{Tr}(*F)_{ab}(*F)^{ab}\, d^4x + \int_{\mathbb{R}^4} \operatorname{Tr} F_{ab}(*F)^{ab}\, d^4x \\
&= S[F] + 2\pi^2 \deg g_\infty.
\end{aligned}$$

The following assertions hold.

1. The equations $\delta T = 0$ and $\delta S = 0$ have the same solutions.

2. The functional $T$ is positive semidefinite and vanishes only on *self-dual connections*, i.e., connections satisfying the *duality equation* discovered by Polyakov and Belavin:

$$F = *F.$$

The solutions to this equation afford global minima to the functional $S$ among the connections with a given value of the degree of the mapping, $\deg g_\infty$.

For $\deg g_\infty = 0$ we have a trivial connection: $F_{ab} \equiv 0$.

For $\deg g_\infty = 1$ there exists a "spherically symmetric" solution to the duality equation (the Belavin–Polyakov–Schwarz–Tyupkin "instanton"), which is given by the following formulas. Define the quantity $\eta_{a\mu\nu}$ by the formulas

$$\begin{aligned}
\eta_{abc} &= \varepsilon_{abc} \quad \text{if } 1 \le a, b, c \le 3, \\
\eta_{ab0} &= \delta_{ab}.
\end{aligned}$$

It is skew-symmetric in $\mu, \nu = 0, 1, 2, 3$ and $a = 1, 2, 3$. We decompose matrices in SU(2) in the basis $e_1, e_2, e_3$ with commutation relations

$$[e_a, e_b] = \varepsilon_{abc} e_c, \qquad X = X^a e_a.$$

Then for any constant $\lambda$ and any point $x_0 \in \mathbb{R}^4$ we obtain the instanton "centered at $x_0$" by the following formulas:

$$A^a_\mu = \frac{2\eta_{a\mu\nu}(x^\nu - x^\nu_0)}{|x - x_0|^2 + \Lambda^2}.$$

For any $N = \deg g_\infty > 1$ we have $N$-instanton solutions

$$A_\mu = -\frac{1}{\rho}\sum_{j=1}^{N} \frac{\lambda_j^2}{|x - x_j|^2} \frac{\partial \omega_j}{\partial x^\mu}\, \omega_j^{-1},$$

where

$$\rho = \sum_{j=1}^{N} \frac{\lambda_j^2}{|x - x_j|^2}, \quad \omega_j = \frac{1}{|x - x_j|}\left((x^0 - x_j^0)\sigma_0 + i(x^k - x_j^k)\sigma_k\right),$$

$\sigma_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\sigma_1$, $\sigma_2$, $\sigma_3$ are the Pauli matrices, the $\lambda_j$ are arbitrary constants, and the $x_j$ are arbitrary points in $\mathbb{R}^4$, $j = 1, \dots, N$.

The duality equations are first-order equations, in contrast with the Yang–Mills equations, which involve the second-order derivatives of the connection. The duality equations for instanton solutions are derived precisely the same way for any four-dimensional Riemannian manifold. We used the sphere $M^4$ only for an elegant interpretation of the integral $D[F]$ as the degree of a mapping and for constructing explicit examples of instantons.

REMARK. There are a number of variational problems, starting from the classical Laplace equation on the plane, which possess the remarkable "instanton phenomenon": their global minima satisfy equations of half the order. We encountered this in the Hodge theory (see above), and this is the case for the Yang–Mills fields on $M^4$. Moreover, there are a number of other important examples as well.

## Exercises to Chapter 15

1. Show that the Maxwell equations in the four-dimensional space-time with metric $g_{ij}$ have the form (15.5):

$$dF = 0, \quad \delta F = \frac{4\pi j}{c},$$

where $j$ is the current, $\delta = *d*$.

2. Prove that the operators of the form

$$R_{ab} - \lambda R g_{ab}$$

are variational derivatives only for $\lambda = 1/2$.

3. Prove that the energy-momentum tensor of the electromagnetic field (14.16) has the form

$$\frac{1}{2} T_{ab} = \frac{1}{\sqrt{-g}} \frac{\delta S}{\delta g^{ab}},$$

where the action $S$ is given by formula (15.4):

$$S = -\frac{1}{16\pi c} \int F_{ab} F_{cd} g^{ac} g^{bd} \sqrt{-g}\, d^4 x, \quad g = \det(g_{ab}).$$

4. Derive the current conservation law

$$\frac{\partial J^a}{\partial x^a} = 0$$

directly from the Euler–Lagrange equations of a massive (scalar or vector) field
$$\frac{\delta S}{\delta\bar{\varphi}}=0,\quad \frac{\delta S}{\delta\varphi}=0 \Longleftrightarrow (\Box+m^2)\varphi=0.$$

5. Show that without the gauge condition
$$\frac{\partial\varphi^a}{\partial x^a}=\frac{\partial\overline{\varphi^a}}{\partial x^a}=0$$
on the vector field the energy $\int T^{00}\,d^3x$ may take negative values.

6. Prove that the image of the two-valued spinor representation $g\colon \mathrm{SO}(3) \to \mathrm{SL}(2,\mathbb{C})$ lies in $\mathrm{SU}(2)$ and the composition of this representation with the projection $\mathrm{SU}(2)\to\mathrm{SU}(2)/\{\pm1\}$ specifies the isomorphism
$$\mathrm{SO}(3)=\mathrm{SU}(2)/\{\pm1\}.$$

7. Consider the space $\mathbb{R}^{1,4}$ with metric $(dx^0)^2-\sum_{k=1}^4(dx^k)^2$ and the hyperboloid in this space specified by the equation
$$(x^0)^2-(x^1)^2-(x^2)^2-(x^3)^2-(x^4)^2=-a^2,\quad a\neq 0.$$
Prove that:

a) This hyperboloid is diffeomorphic to the Cartesian product $\mathbb{R}\times S^3$ and the induced metric can be reduced to the form
$$dt^2-a^2\cosh^2\Bigl(\frac{t}{a}\Bigr)(d\chi^2+\sin^2\chi(d\theta^2+\sin^2\theta d\varphi^2)),$$
where $\chi,\theta,\varphi$ are the angular coordinates on the three-dimensional unit sphere.

b) The metric on the hyperboloid satisfies the Einstein equations in vacuum with nonzero cosmological term $\Lambda\neq 0$ (this space is called the *de Sitter space*).

c) In a de Sitter half-space (for example, in the domain $x^0+x^1>0$) the metric in appropriate coordinates reduces to the static form
$$\Bigl(1-\frac{r^2}{R_\infty^2}\Bigr)dt^2-\Bigl(1-\frac{r^2}{R_\infty^2}\Bigr)^{-1}dr^2-r^2(d\theta^2+\sin^2\theta\,d\varphi^2),\quad R_\infty=\mathrm{const}.$$

8. Let $M_K^3$ be a space of constant sectional curvature $K=0,\pm1$. Its metric has the form
$$dl_K^2=\begin{cases} d\chi^2+\sin^2\chi\,d\Omega^2 & \text{as } K=1 \quad (\text{for } S^3),\\ d\chi^2+\chi^2\,d\Omega^2 & \text{as } K=0 \quad (\text{for } \mathbb{R}^3),\\ d\chi^2+\cosh^2\chi\,d\Omega^2 & \text{as } K=-1 \quad (\text{for } L^3),\end{cases}$$
where $d\Omega^2=d\theta^2+\sin^2\theta\,d\varphi^2$ is the metric on the two-dimensional unit sphere. Let the metric in the space $\mathbb{R}\times M_K^3$ be of the form
$$c^2\,dt^2-a^2(t)\,dl_K^2.$$

Solutions to the Einstein equation in such a class of metrics are called *Friedmann cosmological models.*

a) Within the class of such metrics find the solutions to the Einstein equation with hydrodynamic energy-momentum tensor $T_{ab} = \operatorname{diag}(\varepsilon, p, p, p)$ and prove that, in particular, the following identity holds:

$$\frac{8\pi G\varepsilon}{3c^4} = H^2 + \frac{K}{a^2},$$

where the quantity

$$H = \frac{\dot{a}}{ca}$$

is called the *Hubble constant.*

b) Prove that for dust-cloud matter, when $p = 0$, the solutions mentioned in part a) satisfy the identity

$$\varepsilon a^3 = \text{const}$$

and the following asymptotics holds as $a \to 0$:

$$\varepsilon = \mu c^2 \sim \frac{c^2}{6\pi G} t^{-2}$$

(here $a(t) \to 0$ as $t \to 0$, and $\mu$ is the mass density).

c) Using the equality $\varepsilon a^3 = M = \text{const}$, integrate the Einstein equation by quadratures:

$$\frac{da}{\sqrt{\frac{\lambda}{a} - Kc}} = dt, \quad \varepsilon a^3 = M, \quad \lambda = \frac{8\pi G M}{3c^3}.$$

In particular, show that for $K = -1$ or $0$, the quantity $a$ increases unboundedly, and for $K = 1$, it tends to zero over a finite time.

9. Prove that the scalar product $\langle\psi, \psi\rangle = \psi^+\gamma^0\psi$ (see 15.2.3) is invariant relative to the spinor representation of the group $O(1,3)$.

10. Prove that under the spinor representation of the group $SO(1,3)$ the quantity $\bar{\psi}\gamma^a\psi$ transforms as a vector, $\bar{\psi}\gamma^a\gamma^b\psi$ as a tensor of rank 2, $\bar{\psi}\gamma^a\gamma^b\gamma^c\psi$ as a tensor of rank 3, and $\bar{\psi}\gamma^5\psi = \bar{\psi}\gamma^0\gamma^1\gamma^2\gamma^3\psi$ as a tensor of rank 4.

11. Prove that the semispinor representations of the group $SO(1,3)$ have no nontrivial invariant scalar products.

12. Prove the following relations between the Clifford algebras:

$$\mathrm{Cl}_{n,0} \otimes \mathrm{Cl}_{0,2} = \mathrm{Cl}_{0,n+2}, \quad \mathrm{Cl}_{0,n} \otimes \mathrm{Cl}_{2,0} = \mathrm{Cl}_{n+2,0}, \quad \mathrm{Cl}_{p,q} \otimes \mathrm{Cl}_{1,1} = \mathrm{Cl}_{p+1,q+1}.$$

13. Prove the following isomorphisms of the Clifford algebras:

| $n$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\mathrm{Cl}_{n,0}$ | $\mathbb{C}$ | $\mathbb{H}$ | $\mathbb{H}\oplus\mathbb{H}$ | $\mathrm{M}(2,\mathbb{H})$ |
| $n$ | 5 | 6 | 7 | 8 |
| $\mathrm{Cl}_{n,0}$ | $\mathrm{M}(4,\mathbb{C})$ | $\mathrm{M}(8,\mathbb{R})$ | $\mathrm{M}(8,\mathbb{R})\oplus\mathrm{M}(8,\mathbb{R})$ | $\mathrm{M}(16,\mathbb{R})$ |

| $n$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\mathrm{Cl}_{0,n}$ | $\mathbb{R}\oplus\mathbb{R}$ | $\mathrm{M}(2,\mathbb{R})$ | $\mathrm{M}(2,\mathbb{C})$ | $\mathrm{M}(2,\mathbb{H})$ |
| $n$ | 5 | 6 | 7 | 8 |
| $\mathrm{Cl}_{0,n}$ | $\mathrm{M}(2,\mathbb{H})\oplus\mathrm{M}(2,\mathbb{H})$ | $\mathrm{M}(4,\mathbb{H})$ | $\mathrm{M}(8,\mathbb{C})$ | $\mathrm{M}(16,\mathbb{R})$ |

14. Prove the following periodicity relations for the Clifford algebras:

$$\mathrm{Cl}_{n+8,0} = \mathrm{Cl}_{n,0}\otimes\mathrm{Cl}_{8,0}, \quad \mathrm{Cl}_{0,n+8} = \mathrm{Cl}_{0,n}\otimes\mathrm{Cl}_{0,8}.$$

15. Prove that infinitesimal gauge transformations have the form

$$\psi(x) \to \psi(x) + B(x)\psi(x) + o(B),$$
$$A_\mu(x) \to A_\mu(x) + \nabla_\mu B(x) + o(B).$$

16. Derive Dirac's equation in the presence of a metric.

17. In the space of spinors $\mathbb{C}^2$, let $\xi = (\xi^0, \xi^1)$, $\eta = (\eta^0, \eta^1)$ be a basis such that $\xi^0\eta^1 - \xi^1\eta^0 = 1$. Show that in the Minkowski space, there is a reference system (tetrad) canonically related to this basis, in which the metric becomes

$$(g_{ik}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

18. Prove that if the curvature form of a connection is identically zero, then the connection is (locally) trivial.

19. Prove the Bianchi identity (15.48).

20. Prove that the curvature form of the Cartan connection (see 10.2.4) with values in the Lie algebra of the affine group has the form

$$F_{\mu\nu} = (R^{\kappa}_{\lambda,\mu\nu}, T^{\lambda}_{\mu\nu}),$$

where $R^{\kappa}_{\lambda,\mu\nu}$ is the curvature tensor of the connection and $T^{\lambda}_{\mu\nu}$ is the torsion tensor.

21. Derive the equations of extremals $\frac{\delta S}{\delta A} = 0$ for the Lagrangian

$$L = -\frac{1}{4}\, g^{\mu\lambda} g^{\nu\kappa} \langle F_{\mu\nu}, F_{\lambda\kappa}\rangle,$$

where the metric $g_{\mu\nu}(x)$ is arbitrary and fixed (exterior field).

# Bibliography


[1] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern geometry: Methods and applications*. Part I. *Geometry of surfaces, groups of transformations, and fields*. Part II. *Geometry and topology of manifolds*. 2nd ed., "Nauka", Moscow, 1986; English transl. of 1st ed., *Modern geometry — Methods and applications*. Parts I and II, Graduate Texts in Mathematics, vol. 93 and 104, Springer-Verlag, New York, 1992 and 1985.

[2] ———, *Modern geometry: Methods of homology theory*, "Nauka", Moscow, 1984; English transl., *Modern geometry — Methods and applications*. Part III, *Introduction to homology theory*, Graduate Texts in Mathematics, vol. 124, Springer-Verlag, New York, 1990.

[3] S. P. Novikov and A. T. Fomenko, *Elements of differential geometry and topology*, "Nauka", Moscow, 1987. (Russian)

[4] S. P. Novikov, A. S. Mishchenko, Yu. P. Solovyev, and A. T. Fomenko, *Problems in geometry: Differential geometry and topology*, Moscow State Univ. Press, Moscow, 1978. (Russian)

[5] S. P. Novikov, *Topology*. I, Current problems in mathematics. Fundamental directions. vol. 12, VINITI, Moscow, 1986; English transl., Encyclopaedia of Math. Sciences, vol. 12, Springer-Verlag, Berlin etc., 1996.

[6] V. I. Arnol'd and S. P. Novikov, eds, *Dynamical Systems*, Modern Problems of Mathematics. Fundamental Directions, vol. 4, VINITI, Moscow, 1985; English transl., Encyclopaedia of Math. Sciences, vol. 4, Springer-Verlag, Berlin etc., 1990.

[7] A. D. Aleksandrov, *Intrinsic geometry of convex surfaces*, Gostekhizdat, Moscow–Leningrad, 1948; English transl., A. D. Alexandrov selected works. Part II., Chapman and Hall/CRC, Boca Raton, FL, 2006.

[8] N. V. Efimov, *Higher geometry*, 5th ed., "Nauka", Moscow, 1971; English transl. of 6th ed., "Mir", Moscow, 1980.

[9] A. V. Pogorelov, *Extrinsic geometry of convex surfaces*, "Nauka", Moscow, 1969; English transl., Transl. Math. Monographs, Vol. 35, Amer. Math. Soc., Providence, RI, 1973.

[10] ———, *Differential geometry*, 6th ed., "Nauka", Moscow, 1974; English transl. of 1st ed., Noordhoff, Groningen, 1959.

[11] L. S. Pontryagin, *Smooth manifolds and their applications in homotopy theory*, 3rd ed., "Nauka", Moscow, 1985; English transl. of 2nd ed., L. S. Pontryagin, *Selected works*. Vol. 3. *Algebraic and differential topology*, Gordon and Breach, New York, 1986.

[12] ———, *Continuous groups*, 4th ed., "Nauka", Moscow, 1984; English transl., L. S. Pontryagin, *Selected works*. Vol. 3. *Algebraic and differential topology*, Gordon and Breach, New York, 1986.

[13] P. K. Rashevskii, *Riemannian geometry and tensor analysis*, 3rd ed., "Nauka", Moscow, 1967. (Russian)

[14] E. R. Rosendorn, *Problems in differential geometry*, "Nauka", Moscow, 1971. (Russian)

[15] I. A. Taimanov, *Lectures on differential geometry*, Inst. for Computer Research, Moscow–Izhevsk, 2002. (Russian)

[16] A. T. Fomenko and D. B. Fuks, *A course in homotopic topology*, "Nauka", Moscow, 1989. (Russian)

[17] M. F. Atiyah, *K-Theory*, Benjamin, New York, 1967.

[18] R. Bott and L. W. Tu, *Differential forms in algebraic topology*, Springer-Verlag, New York–Heidelberg–Berlin, 1982.

[19] D. Hilbert and S. Cohn-Vossen, *Geometry and the imagination*, Chelsea Publishing Company, New York, NY, 1952.

[20] D. Gromoll, W. Klingenberg, and W. Meyer, *Riemannsche Geometrie im Grossen*, Lecture Notes in Math., Vol. 55, Springer-Verlag, Berlin–New York, 1968.

[21] H. Seifert and W. Threlfall, *A textbook of topology*, Pure and Appl. Math., vol. 89, Academic Press, New York–London, 1980.

[22] H. Seifert and W. Threlfall, *Variationsrechnung im Grossen*, Teubner, Leipzig, 1938.

[23] S. Lefschetz, *Algebraic topology*, Amer. Math. Soc., New York, 1942.

[24] J. W. Milnor, *Morse theory*, Ann. Math. Studies, no. 51, Princeton Univ. Press, Princeton, NJ, 1963.

[25] ———, *Lectures on the h-cobordism theorem*, Princeton Univ. Press, Princeton, NJ, 1965.

[26] ———, *Singular points of complex hypersurfaces*, Ann. Math. Studies, No. 61, Princeton Univ. Press and the Univ. of Tokyo Press, Princeton, NJ, 1968.

[27] K. Nomizu, *Lie groups and differential geometry*, Math. Soc. Japan, 1956.

[28] J. P. Serre, *Lie algebras and Lie groups*, W. A. Benjamin, New York–Amsterdam, 1965.

[29] G. Springer, *Introduction to Riemann surfaces*, AMS-Chelsea Publishing, Amer. Math. Soc., Providence, RI, 1981.

[30] N. Steenrod, *The topology of fibre bundles*, Princeton Univ. Press, Princeton, NJ, 1951.

[31] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Graduate Studies in Math., Vol. 34, Amer. Math. Soc., Providence, RI, 2001.

[32] S. S. Chern, *Complex manifolds*, Textos de Matematica, No. 5, Inst. de Fisica e Matematica, Univ. do Recife, 1959.

[33] D. McDuff and D. Salamon, *Introduction to symplectic topology*, Clarendon Press, Oxford, 1995.

[34] V. I. Arnol'd, *Ordinary differential equations*, 3rd ed., "Nauka", Moscow, 1984; English transl., Springer-Verlag, Berlin, 1992.

[35] ———, *Additional chapters of the theory of ordinary differential equations*, "Nauka", Moscow, 1978; English transl., *Geometrical methods in the theory of ordinary differential equations*, Grundlehren der Mathematischen Wissenschaften, vol. 250, Springer-Verlag, New York-Berlin, 1983.

[36] ———, *Mathematical methods of classical mechanics*, 3rd ed., "Nauka", Moscow, 1989; English transl., Graduate Texts in Math., vol. 60, Springer-Verlag, New York, 1989.

[37] V. V. Golubev, *Lectures on integration of the equations of motion of a rigid body about a fixed point*, Gostehizdat, Moscow, 1954; English transl., Israel Program for Scientific translations, 1960.

[38] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, McGraw-Hill, New York, 1955.

[39] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., "Nauka", Moscow, 1973; English transl., *Course of theoretical physics.* Vol. 1. *Mechanics.* Pergamon Press, Oxford–New York–Toronto, 1976.

[40] L. S. Pontryagin, *Ordinary differential equations*, 4th ed., "Nauka", Moscow, 1974. (Russian)

[41] A. A. Abrikosov, *Fundamentals of the theory of metals*, "Nauka", Moscow, 1987. (Russian)

[42] A. I. Akhiezer and V. B. Berestetskii, *Quantum electrodynamics*, 3rd ed., "Nauka", Moscow, 1969; English transl. of 2nd ed., Wiley, New York–London–Sydney, 1965.

[43] N. N. Bogolyubov and D. V. Shirkov, *Introduction to quantum field theory*, 4th ed., "Nauka", Moscow, 1984. (Russian)

[44] O. I. Bogoyavlensky, *Methods in the qualitative theory of dynamical systems in astrophysics and gas dynamics*, "Nauka", Moscow, 1980; English transl., Springer-Verlag, Berlin, 1985.

[45] J. D. Bjorken and S. D. Drell, *Relativistic quantum fields*, McGraw-Hill, New York, 1965.

[46] Ya. B. Zel'dovich and I. D. Novikov, *Structure and evolution of the Universe*, "Nauka", Moscow, 1975; English transl., Univ. of Chicago Press, 1983.

[47] L. D. Landau and E. M. Lifschitz, *Theoretical physics*, Vol. II, *Theory of fields*, "Nauka", Moscow, 1973. English transl., Pergamon Press, Oxford–New York–Toronto, 1975.

[48] ———, *The mechanics of continuous media*, "Nauka", Moscow, 1973. English transl., *Course of theoretical physics*, vol. 6, *Fluid mechanics*, and vol. 7, *Theory of elasticity*, Pergamon Press, London–Paris–Frankfurt; Addison-Wesley, Reading, MA, 1959.

[49] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman, San Francisco, 1973.

[50] R. E. Peierls, *Quantum theory of solids*, Oxford Univ. Press, Oxford, 2001.

[51] L. I. Sedov, *Mechanics of continuous media*, 4th ed., "Nauka", Moscow, 1983–1984; English transl., vol. 1, 2, Ser. Theoret. and Appl. Mechanics, 4. World Scientific, River Edge, NJ, 1997.

[52] A. A. Slavnov and L. D. Faddeev, *Introduction to the quantum theory of gauge fields*, 2nd ed., "Nauka", Moscow, 1988; English transl., *Gauge fields. Introduction to quantum theory*, Frontiers in Physics, 83, Addison-Wesley, Redwood City, CA, 1991.

[53] S. P. Novikov, ed., *Theory of solitons*, "Nauka", Moscow, 1979. (Russian)

[54] A. Wintner, *The analytical foundations of celestial mechanics*, Princeton Math. Series, Vol. 5, Princeton Univ. Press, Princeton, NJ, 1941.

[55] R. P. Feynman, R. B. Leighton, and M Sands, *The Feynman lectures on physics*, vols. 1–3, Addison-Wesley, Reading, MA, 1963–1965.

[56] A. M. Polyakov, *Gauge fields and strings*, Harwood Academic Publishers, Chur, 1987.

[57] B. A. Dubrovin and S. P. Novikov, *Hydrodynamics of weakly deformed soliton lattices. Differential geometry and Hamiltonian theory*, Uspekhi Mat. Nauk **44** (1989), 29–98; English transl., Russian Math. Surveys **44** (1989), 35–124.

[58] Le Ty Kuok Tkhang, S. A. Piunikhin, and V. A. Sadov, *The geometry of quasicrystals*, Uspekhi Mat. Nauk **48** (1993), 41–102; English transl., Russian Math. Surveys **48** (1993), 37–100.

[59] S. P. Novikov, *The Hamiltonian formalism and a multivalued analogue of Morse theory*, Uspekhi Mat. Nauk **37** (1982), 3–49; English transl., Russian Math. Surveys **37** (1982), 1–56.

[60] S. P. Novikov and A. Ya. Maltsev, *Topological phenomena in normal metals*, Uspekhi Fiz. Nauk **168** (1998), 249–258. (Russian)

[61] L. D. Faddeev, *The energy problem in Einstein's theory of gravitation*, Uspekhi Fiz. Nauk **136** (1982), 435–457. (Russian)

[62] I. A. Taimanov, *Modified Novikov–Veselov equation and differential geometry of surfaces*, Transl. Amer. Math. Soc., Ser. 2, vol. 179, Amer. Math. Soc., Providence, RI, 1997, pp. 133–151.

# Index

## 美国数学会经典影印系列

| | | |
|---|---|---|
| 1 | **Lars V. Ahlfors**, Lectures on Quasiconformal Mappings, Second Edition | 9787040470109 |
| 2 | **Dmitri Burago**, **Yuri Burago**, **Sergei Ivanov**, A Course in Metric Geometry | 9787040469080 |
| 3 | **Tobias Holck Colding**, **William P. Minicozzi II**, A Course in Minimal Surfaces | 9787040469110 |
| 4 | **Javier Duoandikoetxea**, Fourier Analysis | 9787040469011 |
| 5 | **John P. D'Angelo**, An Introduction to Complex Analysis and Geometry | 9787040469981 |
| 6 | **Y. Eliashberg**, **N. Mishachev**, Introduction to the *h*-Principle | 9787040469028 |
| 7 | **Lawrence C. Evans**, Partial Differential Equations, Second Edition | 9787040469356 |
| 8 | **Robert E. Greene**, **Steven G. Krantz**, Function Theory of One Complex Variable, Third Edition | 9787040469073 |
| 9 | **Thomas A. Ivey**, **J. M. Landsberg**, Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems | 9787040469172 |
| 10 | **Jens Carsten Jantzen**, Representations of Algebraic Groups, Second Edition | 9787040470086 |
| 11 | **A. A. Kirillov**, Lectures on the Orbit Method | 9787040469103 |
| 12 | **Jean-Marie De Koninck**, **Armel Mercier**, 1001 Problems in Classical Number Theory | 9787040469998 |
| 13 | **Peter D. Lax**, **Lawrence Zalcman**, Complex Proofs of Real Theorems | 9787040470000 |
| 14 | **David A. Levin**, **Yuval Peres**, **Elizabeth L. Wilmer**, Markov Chains and Mixing Times | 9787040469943 |
| 15 | **Dusa McDuff**, **Dietmar Salamon**, *J*-holomorphic Curves and Symplectic Topology | 9787040469936 |
| 16 | **John von Neumann**, Invariant Measures | 9787040469974 |
| 17 | **R. Clark Robinson**, An Introduction to Dynamical Systems: Continuous and Discrete, Second Edition | 9787040470093 |
| 18 | **Terence Tao**, An Epsilon of Room, I: Real Analysis: pages from year three of a mathematical blog | 9787040469004 |
| 19 | **Terence Tao**, An Epsilon of Room, II: pages from year three of a mathematical blog | 9787040468991 |
| 20 | **Terence Tao**, An Introduction to Measure Theory | 9787040469059 |
| 21 | **Terence Tao**, Higher Order Fourier Analysis | 9787040469097 |
| 22 | **Terence Tao**, Poincaré's Legacies, Part I: pages from year two of a mathematical blog | 9787040469950 |
| 23 | **Terence Tao**, Poincaré's Legacies, Part II: pages from year two of a mathematical blog | 9787040469967 |
| 24 | **Cédric Villani**, Topics in Optimal Transportation | 9787040469219 |
| 25 | **R. J. Williams**, Introduction to the Mathematics of Finance | 9787040469127 |
| 26 | **T. Y. Lam**, Introduction to Quadratic Forms over Fields | 9787040469196 |

| | | |
|---|---|---|
| 27 | **Jens Carsten Jantzen**, Lectures on Quantum Groups | 9 787040 469141 |
| 28 | **Henryk Iwaniec**, Topics in Classical Automorphic Forms | 9 787040 469134 |
| 29 | **Sigurdur Helgason**, Differential Geometry, Lie Groups, and Symmetric Spaces | 9 787040 469165 |
| 30 | **John B.Conway**, A Course in Operator Theory | 9 787040 469158 |
| 31 | **James E. Humphreys**, Representations of Semisimple Lie Algebras in the BGG Category $\mathcal{O}$ | 9 787040 468984 |
| 32 | **Nathanial P. Brown**, **Narutaka Ozawa**, C*-Algebras and Finite-Dimensional Approximations | 9 787040 469325 |
| 33 | **Hiraku Nakajima**, Lectures on Hilbert Schemes of Points on Surfaces | 9 787040 501216 |
| 34 | **S. P. Novikov**, **I. A. Taimanov**, **Translated by Dmitry Chibisov**, Modern Geometric Structures and Fields | 9 787040 469189 |
| 35 | **Luis Caffarelli**, **Sandro Salsa**, A Geometric Approach to Free Boundary Problems | 9 787040 469202 |
| 36 | **Paul H. Rabinowitz**, Minimax Methods in Critical Point Theory with Applications to Differential Equations | 9 787040 502299 |
| 37 | **Fan R. K. Chung**, Spectral Graph Theory | 9 787040 502305 |
| 38 | **Susan Montgomery**, Hopf Algebras and Their Actions on Rings | 9 787040 502312 |
| 39 | **C. T. C. Wall**, **Edited by A. A. Ranicki**, Surgery on Compact Manifolds, Second Edition | 9 787040 502329 |
| 40 | **Frank Sottile**, Real Solutions to Equations from Geometry | 9 787040 501513 |
| 41 | **Bernd Sturmfels**, Gröbner Bases and Convex Polytopes | 9 787040 503081 |
| 42 | **Terence Tao**, Nonlinear Dispersive Equations: Local and Global Analysis | 9 787040 503050 |
| 43 | **David A. Cox**, **John B. Little**, **Henry K. Schenck**, Toric Varieties | 9 787040 503098 |
| 44 | **Luca Capogna**, **Carlos E. Kenig**, **Loredana Lanzani**, Harmonic Measure: Geometric and Analytic Points of View | 9 787040 503074 |
| 45 | **Luis A. Caffarelli**, **Xavier Cabré**, Fully Nonlinear Elliptic Equations | 9 787040 503067 |